Universes in Type Theory Part I – Inaccessibles and Mahlo

Anton Setzer[*]

July 8, 2005

Abstract

We give an overview over universes in Martin-Löf type theory and consider the following universe constructions: a simple universe, E. Palmgren's super universe and the Mahlo universe. We then introduce models for these theories in extensions of Kripke-Platek set theory having the same proof theoretic strength. The extensions of Kripke-Platek set theory used formalise the existence of a recursively inaccessible ordinal, a recursively hyper-inaccessible ordinal, and a recursively Mahlo ordinal. Using these models we determine upper bounds for the proof theoretic strength of the theories in questions. In case of simple universes and the Mahlo universe, these bounds have been shown by the author to be sharp. This article is an overview over the main techniques in developing these models, full details will be presented in a series of future articles.

1 Introduction

This article presents some results of a research program with the goal of determining as strong as possible predicatively justified extensions of Martin-Löf type theory (MLTT) and to determine their precise proof theoretic strength. We see three main reasons for following this research program:

1. The goal is to develop type theoretic analogues of the theories analysed in proof theory. This way we hope to make the rather abstract and technically difficult results from proof theory more accessible to the general audience, and we hope as well to give some computational meaning to those results. This will be particularly important in case of the Π_3-reflecting universe (to be presented in the followup article [32]) which was developed from Rathjen's ordinal notation system for $\mathrm{KP} + (\Pi_3 - \mathrm{refl})$ (Kripke-Platek set theory extended by the principle of Π_3-reflection).

[*]Department of Computer Science, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, UK, Email: `a.g.setzer@swan.ac.uk`, `http://www.cs.swan.ac.uk/~csetzer/`, Tel: +44 1792 513368, Fax: +44 1792 295651. Supported by Nuffield Foundation, grant ref. NAL/00303/G and EPSRC grant GR/S30450/01.

2. One can consider predicatively justified extensions of MLTT as safe theories, namely theories with a philosophical argument which justifies the validity of everything shown in those theories. This philosophical argument is given by Martin-Löf's meaning explanations. The proofs of the lower bounds show that these extensions of MLTT show the consistency of corresponding extensions of Kripke-Platek set theory, or more precisely of any approximation of it. If one accepts meaning explanations as a philosophical argument for the validity of statements shown in MLTT, one obtains in this way a consistency proof for strong extensions of Kripke-Platek set theory, which are strong enough to prove a large portion of mathematical theories (see the results in reverse mathematics, e.g. Simpson's book [33], which show that rarely more than the strength of $(\Pi^1_1 - \mathrm{CA})_0$ is needed; one exception seems to be the graph minor theorem). These results are in line with a revised Hilbert's program. Hilbert's original program was to prove the consistency of axiom systems for formalising mathematical proofs by finitary arguments. By Gödel's incompleteness theorem we know that this program cannot be carried out. MLTT can be considered as one replacement for finitary methods in a revised Hilbert's program, and our research demonstrates that MLTT can well be used for this purpose.

   We should note however that meaning explanations haven't been worked out yet in the case of the Mahlo universe – the author himself doesn't have a sufficient background in philosophy to fill in this gap.

3. We hope as well that the new sets developed can be used as data structures in general computing. This hope has been fulfilled in case of the Mahlo universe. The data type of inductive recursive definitions used in the closed formalisation of inductive-recursive definitions developed by P. Dybjer and the author (see [10, 12, 11, 13]) has a similar character as the Mahlo universe. Variants of this data type have been used in [9, 7] in the area of generic or generative programming with the goal of writing programs which not only use data but as well the buildup of data structures given to them in order to compute new data structures and data. We hope that the Π_3-reflecting universe, which will be developed in the followup article [32], will give rise to similar new data structures.

We should note that whether the Mahlo universe and extensions of it presented in the follow up article [32] are actually predicatively justifiable is still a matter of a debate. However, even if there is a debate on this, it seems at least at the moment to be unlikely that a better construction can be found which avoids the problem in this case. For instance Erik Palmgren's higher order universes [20] look at first sight more directly predicatively justifiable – but a closer look (it took the author a long time to actually discover this) reveals that in Palmgren's construction Set has a similar character as a Mahlo universe: The introduction rule allows to introduce new elements of Set assuming an arbitrary function from families of Set into families of Set (the new element will be a universe closed under this function). This is essentially the same as the introduction rule for the Mahlo universe.

Follow up articles. The original goal when writing this article was to present the Π_3-reflecting universe. It turned out that much more space is needed to present it, and therefore the presentation of the Π_3-reflecting universe will be given in a planned followup article [32]. Even so the current article is rather long, we won't have room to present the models in full detail. We plan to write a series of articles, in which the models will be worked out in full detail. These future articles will then set up a better infrastructure for writing future articles on model constructions for MLTT. So the current article serves as an overview article, which presents the basic techniques for developing models of type theory suitable for determining upper bounds for the proof theoretic strength. It might be that this article is more accessible than the more detailed technical articles to follow.

Content. The structure of this article is as follows: In Sect. 2 we briefly develop the small and large logical framework and the basic set constructions (i.e. the sets N_n, N and the sets formed using $+$, Σ, Π, W, Id). We then briefly introduce Kripke-Platek set theory. Then we develop the basic principles for defining models of type theory for the basic set constructions and the small logical framework. In Sect. 3 we introduce the rules for universes and the theory $\mathrm{ML}_{\mathrm{W}}\mathrm{U}$. Furthermore we determine the principles for introducing nested universes. We will see that there are two ways of defining subuniverses: recursive and inductive subuniverses. We then develop the main concepts for introducing models for universes and develop in particular a model of $\mathrm{ML}_{\mathrm{W}}\mathrm{U}$. In Sect. 4 we introduce Erik Palmgren's super universe and develop a model of it in a corresponding extension of Kripke-Platek set theory. Finally, in Sect. 5 we introduce the Mahlo universe and determine a model of it. For all three universe constructions mentioned we obtain proof theoretic upper bounds for the strength of the theories in question.

Related results. The first model of MLTT developed for proof theoretic purposes was the model of MLTT with one universe but no W-type, developed by P. Aczel ([1]). Rathjen and Griffor have analysed the strength of MLTT with finitely many universes, W-type, and of several variants of it in [14]. Erik Palmgren has introduced the super universe and later higher order universes (the latter ones are conjectured to reach the strength of KPM), which are best presented in [20]. There is a rich literature on PER models of MLTT, an overview is given in M. Hofmann's article [15]. Note that the emphasis of this article is in developing models in order to determine upper bounds for the proof-theoretic strength of those theories, which means as well to develop models using minimal strength in the Meta-theory. In this research programme we are following the steps of ordinal theoretic proof theory, in which the major recent steps were done by M. Rathjen, who analysed the theories KPM ([22]), $\mathrm{KP} + (\Pi_3 - \mathrm{refl})$ [23] and $\Pi^1_2{-}\mathrm{CA}{+}\mathrm{BI}$ ([24]). Similar steps have been taken by T. Arai ([3, 2, 4, 5];

there are as well preprints of Arai covering Π_n-reflection, Σ_1-stability and Π_1-collection).

This research benefitted very much from intensive discussions with T. Arai, U. Berger, P. Martin-Löf, E. Palmgren, and M. Rathjen.

2 Basic Martin-Löf Type Theory

In this section we will briefly introduce the version of Martin-Löf type theory (MLTT) used in this article. We will require some knowledge about MLTT. The reader not familiar with it might refer to [17, 18, 19, 21] or Chapter 11 of [34] (the latter deviates in its description of intensional type theory from the standard versions of MLTT).

2.1 The Small Logical Framework

The role of the logical framework in this article. In this article we will, when formally introducing theories and analysing them, not make use of the logical framework. There are two reasons for it:

On one hand, the author has at the moment conceptual problems with the logical framework. The problem manifests itself particularly when considering meaning explanations. Meaning explanations for the logical framework don't seem to have been worked out fully at present, whereas the concept of meaning explanations for type theory without the logical framework seems to be well understood. One should note however that Per Martin-Löf seems to have a clear understanding of meaning explanations for the logical framework, and has given talks on this topic.

On the other hand, a full treatment of the logical framework causes at the moment still problems when determining proof theoretic bounds. Our techniques for modelling type theories at present don't allow to model it directly without using more strength than is actually needed. Therefore, with our techniques we won't be able to obtain precise proof theoretic bounds. In order to avoid this, it seems to be necessary to first eliminate the use of the logical framework, and then to use the techniques used in this article. Martin Hofmann has shown in Sect. 4.3 of [15] that type theory with the logical framework is conservative over type theory without it, which would give the desired reduction. However, we haven't had time yet to study this result in detail yet, in order to make sure that it can be applied to our setting.

For these two reasons all theories presented in the following will only make use of the restriction of the logical framework to set, which we call the small logical framework. However, in order to explain the heuristics of our constructions, it is sometimes useful to make use of the full logical framework. We will do so for heuristic purposes only – when the formal theories are introduced, we will not make use of it.

The small logical framework. The small logical framework contains the dependent function set and product as in the the full logical framework, but limited to sets. So we have, under the assumptions $A : \mathrm{Set}$, $x : A \Rightarrow B : \mathrm{Set}$ the following set constructions:

- The dependent function set $(x : A) \to B : \mathrm{Set}$.
  - The introduction rule expresses that we can form $(x : A)t : (x : A) \to B$ provided $x : A \Rightarrow t : B$.
  - The elimination rule expresses that we can apply $f : (x : A) \to B$ to $a : A$ and obtain $f(a) : B[x := a]$.
  - We have as equality rule β-equality, i.e. $((x : A)t)s = t[x := s]$.
  - Additionally we have the η-rule: if $f : (x : A) \to B$ then $f = (x : A)f(x)$.
  - Furthermore we have equality rules of the formation rule (the rule which forms $(x : A) \to B : \mathrm{Set}$, provided $A : \mathrm{Set}$ and $x : A \Rightarrow B : \mathrm{Set}$), the introduction rule and the elimination rule:
    * The equality version of the formation rule expresses that if $A = A' : \mathrm{Set}$, $x : A \Rightarrow B = B' : \mathrm{Set}$, then $(x : A) \to B = (x : A') \to B' : \mathrm{Set}$.
    * The equality version of the introduction rule expresses that if $x : A \Rightarrow t = t' : B$ then $(x : A)t = (x : A)t'$ (that's the ξ-rule).
    * The equality version of the elimination rule expresses that if $s = s' : (x : A) \to B$ and $t = t'$ then $s(t) = s'(t')$.
    * The principle of forming the equality versions of the formation, introduction and elimination rules from the standard (non-equality) rules is a straightforward principle. Therefore the convention is that when we introduce rules in the following, we silently introduce as well the equality versions as well. An exception are the equality rules, which don't have an equality version.
    * α-equivalent terms are considered to be identical, therefore there is no explicit rule for α-equality. This extends to judgements as well, i.e. $x : A, y : B \Rightarrow C : \mathrm{Set}$ and $u : A, v : B[x := u] \Rightarrow C[x := u, y := v] : \mathrm{Set}$ are considered as the same judgement.
  - We mention the following abbreviations:
    * We write $f(a, b, c)$ instead of $f(a)(b)(c)$, similarly for longer applications.
    * We usually omit the type A in $(x : A)t$ and write $(x)t$ instead.
    * $(x, y)t := (x)(y)t$, similarly for longer expressions.
    * We write $(x, y : A) \to B$ for $(x : A) \to (x : B) \to B$, $(x : A, y : B) \to C$ for $(x : A) \to (y : B) \to C$. Similar abbreviations are to be understood in the same way.
    * If $f : A \to B$ and $g : B \to C$ then $g \circ f := (x)g(f(x))$.

- The dependent product $(x : A) \times B : \mathrm{Set}$.
  - The introduction rule allows to form $\langle a, b\rangle$ for $a : A$ and $b : B[x := a]$.
  - The elimination rule forms the projections of an element $c : (x : A) \times B : \mathrm{Set}$, written as $\pi_0(c) : A$ and $\pi_1(c) : B[x := \pi_0(c)]$.
  - The equality rule expresses that $\pi_0(\langle a, b\rangle) = a$ and $\pi_1(\langle a, b\rangle) = b$.
  - We have as well the η-rule associated with it, so if $c : (x : A) \times B$ then $c = \langle \pi_0(c), \pi_1(c)\rangle$.

We introduce here as well $A \to B := (x : A) \to B$ and $A \times B := (x : A) \times B$ for some fresh variable x.

Definition 2.1. *By the rules of the small logical framework we mean the structural rules of type theory and the rules for the dependent function set and the dependent product restricted to* Set.

2.2 The Full Logical Framework

The (full) logical framework (which is never part of the official type theories in this article), is obtained in the following way: Apart from the type sets, which is the highest type in the version excluding the logical framework, one has as well a type of large sets Type. Type contains Set and every element of Set. So we have the rules

$$\mathrm{Set} : \mathrm{Type} \qquad \frac{A : \mathrm{Set}}{A : \mathrm{Type}} \qquad \frac{A = B : \mathrm{Set}}{A = B : \mathrm{Type}}$$

Apart from Set, Type will as well be closed under the dependent function type and the dependent product. So if $A : \mathrm{Type}$ and $x : A \Rightarrow B : \mathrm{Type}$, we have that

- $(x : A) \to B : \mathrm{Type}$,
- $(x : A) \times B : \mathrm{Type}$,

with essentially the same rules as the small logical framework, except that one refers to Type instead of Set.

2.3 The Basic Set Constructions

The basic set constructions are the following sets (or more precisely principles for forming sets) and their corresponding rules:

- The finite sets $\mathrm{N}_n : \mathrm{Set}$, where $n \in \{0, 1, \ldots, \})$; note that here n is not an internal natural number inside type theory, but exists on the Meta-level. The introduction rules are $\mathrm{A}_i^n : \mathrm{N}_n$ for $i = 0, \ldots, n-1$ (where i are again numbers on the Meta-level).
- The set N of natural numbers. The introduction rules are $0 : \mathrm{N}$ and $\mathrm{S} : \mathrm{N} \to \mathrm{N}$.

- The disjoint union $A + B$ of two sets A, B, with introduction rules $\mathrm{inl} : A \to (A + B)$ and $\mathrm{inr} : B \to (A + B)$.
- For $A : \mathrm{Set}$ and $x : A \Rightarrow B : \mathrm{Set}$ we have the following sets:
  - The Π-set $\Pi x : A.B$. The introduction rule is $\lambda : ((x : A) \to B) \to \Pi x : A.B$.
    One writes $\lambda x : A.t$ for $\lambda((x : A)t)$.
  - The Σ-set $\Sigma x : A.B$. The introduction rule is $\mathrm{p} : (a : A) \to (b : B[x := a]) \to \Sigma x : A.B$.
    (The differences between $\Pi x : A.B$, $\Sigma x : A.B$, and $(x : A) \to B$, $(x : A) \times B$ will be explained below).
  - The W set $\mathrm{W} x : A.B$, which is the set of well-founded trees with branching degrees $B[x := a]$ for $a : A$. The introduction rule is $\sup : (a : A, b : B[x := a] \to \mathrm{W} x : A.B) \to \mathrm{W} x : A.B$.
- The intensional identity set $\mathrm{Id}(A, a, b)$ for $A : \mathrm{Set}$, $a : A$, $b : A$. The introduction rule is $\mathrm{refl}_A : (a : A) \to \mathrm{Id}(A, a, a)$.
- For all of the above set constructions, the elimination rules express that the above sets are the least sets introduced by these constructors. This will be for instance in case of N primitive recursion into arbitrary sets (which might depend on the element $n : \mathrm{N}$ we are eliminating), in case of W induction over those trees, and in case of $A + B$ case distinction on whether $ab : A + B$ is of the form $\mathrm{inl}(a)$ or $\mathrm{inr}(b)$.
- Furthermore, we have the standard equality rules for the above mentioned sets.

The main difference between the dependent function set and product and the sets $\Pi x : A.B$, $\Sigma x.B$ is that we have the η-rule for the constructions of the logical framework, but not for Π and Σ. The conceptual reason for this becomes clear when considering inductive-recursive definitions. All the above constructions are inductive-recursive (in case of $\mathrm{Id}(A, a, b)$ general indexed inductive-recursive definitions) as introduced originally due to P. Dybjer with a formalisation using finitely many rules by P. Dybjer and the author; see [10, 8, 12, 11, 13]). Indexed inductive-recursive definitions allow to introduce all sets in MLTT by determining their introduction rules (there are of course restrictions on which introduction rules are allowed). The elimination and equality rules are then derived automatically. There is no η-rule involved in this schema, which would be unnatural in general. For instance, it does not make much sense to consider an η-rule for N. Therefore it is natural to exclude the η-rule from Π and Σ, and to have separate logical framework versions which contain the η-rule. In this article, the difference between the logical framework set constructions and the Π- and Σ-set won't play a big rôle.

Our models will admit as well the addition of an extensional equality set to the type theory. Since lower bounds will be obtained using intensional equality only, it will follow that the proof theoretic strength of the type theories under consideration with intensional equality and with extensional equality coincides.

Definition 2.2. *By the basic set constructions in type theory we mean the above mentioned constructions* N_n, N, $+$, Σ, Π, W, Id *for forming sets and the corresponding formation/introduction/elimination/equality rules.*

2.4 Kripke-Platek Set Theory

In this article, we will develop models of the type theories considered in versions of Kripke-Platek set theory. This will allow us to determine upper bounds for the theories in question.

Kripke-Platek set theory (KP) is a weak version of set theory, based on classical logic. The "bible" of KP is the book by Barwise [6]. KP can be used in order to develop most concepts in generalised recursion theory, as demonstrated in that book. KP has been pioneered by Jäger [16] as a reference theory for proof theoretic studies. For many theories, there exists a version of KP of equal strength, and often one can determine upper bounds for theories by determining upper bounds for that variant for KP and then by modelling the original theory in that variant. This is as well the approach taken in this article.

We don't have room to introduce KP in full (we highly recommend the reader not familiar with it to study the first chapters of [6]). We briefly repeat here its axioms (see p. 10 of [6]; the theory presented in [6] adds to KP as well urelemente, and is therefore called KPU for KP plus urelemente; the versions of KP used in this article don't have urelemente):

(Extensionality)	$\forall x, y.(\forall z.z \in x \leftrightarrow z \in y) \rightarrow x = y$
(Foundation)	$\forall \vec{z}.(\forall x.(\forall y \in x.\varphi(y, \vec{z})) \rightarrow \varphi(x, \vec{z})) \rightarrow \forall x.\varphi(x, \vec{z})$ where $y \notin \mathrm{FV}(\varphi(x, \vec{z}))$
(Pair)	$\forall x, y.\exists z.x \in z \wedge y \in z$
(Union)	$\forall x.\exists y.\forall z \in x.\forall u \in z.u \in y$
(Δ_0-Separation)	$\forall \vec{x}, y.\exists z.\forall u.(u \in z \leftrightarrow (u \in y \wedge \varphi(x, \vec{x})))$ where $\varphi(x, \vec{x})$ is Δ_0, $z \notin \mathrm{FV}(\varphi(x, \vec{x}))$
(Δ_0-Collection)	$\forall \vec{x}, y.(\forall z \in y.\exists u.\varphi(z, u, \vec{x})) \rightarrow \exists v.\forall z \in y.\exists u \in v.\varphi(z, u, \vec{x})$ where $\varphi(z, u, \vec{x})$ is Δ_0, $v \notin \mathrm{FV}(\varphi(z, u, \vec{x}))$

The versions of KP used in this article will always be augmented by the axiom of infinity

(Infinity) $\exists x.\emptyset \in x \wedge \forall y \in x.y \cup \{y\} \in x$,

where $\emptyset \in x$ and $y \cup \{y\} \in x$ are to be understood in the usual way. Let KPω be the theory KP plus the axiom of infinity. When forming extensions of KPω, one adds a predicate $\mathrm{Ad}(x)$ for "x is an admissible containing ω", following the approach taken by Jäger (e.g. [16]). Here an admissible is a transitive inner model of KP, and by admissible $> \omega$ we mean an admissible containing ω, i.e.

a transitive inner model of KPω. So one has the following additional axioms (trans(x) expresses that x transitive):

(Ad.1)	$\forall x.\mathrm{Ad}(x) \to \mathrm{trans}(x)$
(Ad.2)	$\forall x, y.\mathrm{Ad}(x) \wedge \mathrm{Ad}(y) \to (x \in y \vee x = y \vee y \in x)$
(Ad.3)	$\forall x, \vec{y}.(\mathrm{Ad}(x) \to \psi^x(\vec{y}))$ for every instance $\psi(\vec{y})$ of (*Pair*), (*Union*), (Δ_0-*Separation*), (Δ_0-*Collection*)

$\psi^x(\vec{y})$ is obtained by replacing all unbounded quantifiers $\forall y$, $\exists y$ occurring in $\psi(\vec{y})$ by $\forall y \in x$, $\exists y \in x$ respectively, but leaving restricted quantifiers as they are (where an unrestricted quantifier is a quantifier not of the form $\forall y \in z$, $\exists y \in z$). The extensions of KPω are then obtained by adding to KPω the above axioms for Ad and axioms expressing that there exists an admissible with sufficiently strong closure properties. The convention in this article is that whenever we introduce an extension of KP, we always add as well the infinity axiom and, if Ad is involved in any of the additional axioms, the axioms *(Ad.1-3)*.

We introduce as well the theory KPl which expresses that there are finitely many admissibles, but that the set theoretic universe itself is not necessarily a model of KP. The standard model of KPl is $\mathrm{a} := \bigcup_{n\in\omega} \mathrm{b}_n$, where b_n is the nth admissible ($n \in \omega$). Note that a is not an admissible. Formally, KPl is obtained by taking KPω without (Δ_0-*Collection*) and adding the axioms *(Ad.1 -3)* and the following axiom *(Lim)*:

(Lim) $\forall x.\exists y.\mathrm{Ad}(y) \wedge x \in y$

When working in extensions of KPω, we will frequently make use of the constructible hierarchy, and refer to the definition of L_α in [6]. We define as well the following:

Definition 2.3. *(a)* $\mathrm{L}_{<\alpha} := \bigcup_{\beta<\alpha} \mathrm{L}_\beta$,

(b) $\mathrm{L}'_\alpha := \mathrm{L}_{\omega\cdot\alpha}$,

(c) $\mathrm{L}'_{<\alpha} := \bigcup_{\beta<\alpha} \mathrm{L}'_\beta$.

(d) An admissible ordinal is an ordinal α s.t. L_α is admissible.

We will often make use of the following essentially trivial remark:

Remark 2.4. *(a) If $b = \{x \in a \mid \varphi(x, a_1, \ldots, a_n)\}$ where φ is Δ_0 and $a, a_1, \ldots, a_n \in \mathrm{L}'_\alpha \cup \{\mathrm{L}'_\alpha\}$, then $b \in \mathrm{L}'_{\alpha+1}$.*

(b) $\alpha \in \mathrm{L}'_{\alpha+1}$.

(c) $\omega \in \mathrm{L}'_2$.

Proof. Let Ord be the class of ordinals. (a): Assumption II.5.2 and Theorem II.6.4 of [6]. (b): One easily can show that $\mathrm{L}_\alpha \cap \mathrm{Ord} \in \mathrm{Ord}$. We show the assertion by induction on α: The cases $\alpha = 0$ and $\alpha \to \alpha + 1$ follow easily or by

IH. In case α limit we have by IH $\alpha \subseteq \mathrm{L}'_{<\alpha} \cap \mathrm{Ord} =: \beta \in \mathrm{L}'_{\alpha+1}$, and therefore $\alpha \in \mathrm{L}'_{\alpha+1}$. (c): $\omega \subseteq \mathrm{L}_{<\omega}$, $\mathrm{L}_{<\omega} \cap \mathrm{Ord} \subseteq \omega$, therefore $\omega := \{x \in \mathrm{L}_{<\omega} \mid \mathrm{Ord}(x)\} \in \mathrm{L}_{\omega+\omega} = \mathrm{L}'_2$.

$\square$

2.5 Models of the Basic Type Theory

As said before, we will develop models of the type theories in question in extensions of KPω. These models will extend the model of MLTT with one universe and the W-type, as worked out in the author's PhD thesis [25] (see as well [27] for a draft version of an article on it, and [26] for an article containing a basic model construction for MLTT with one Mahlo universe).

Basic principles of the model construction. The basic idea of all our models is that we define a PER model, i.e. a model in which sets are interpreted as partial equivalence relations on a set of basic objects, where in this article the set of basic objects will be a set of terms, which is a slightly simplified version of the set of terms occurring in the type theory. Let ML$'$ the type theory for which we want to determine an upper bound for its proof theoretic strength. We will model ML$'$ in an extension KP$'$ of KPω. KP$'$ will be defined in such a way that it has the same expected proof theoretic strength. Once we have modelled ML$'$ in KP$'$ one will be able to conclude (this step requires some techniques which can be found in [25]) that the proof theoretic strength $|\mathrm{ML}'|$ of ML$'$ is less than or equal the proof theoretic strength $|\mathrm{KP}'|$ of KP$'$, i.e. $|\mathrm{ML}'| \leq |\mathrm{KP}'|$.

The soundness theorem for the model of type theory will be a Meta-theorem (otherwise we would obtain $|\mathrm{ML}'| < |\mathrm{KP}'|$): For Meta-every judgement $\Gamma \Rightarrow \theta$ s.t. $\mathrm{ML}' \vdash \Gamma \Rightarrow \theta$ we have that KP$'$ proves the correctness statement associated with $\Gamma \Rightarrow \theta$. But ML$'$ doesn't prove that for all $\Gamma \Rightarrow \theta$ the correctness statement holds. In fact, ML$'$ won't even express such a general correctness statement.

Let Term be the underlying set of terms occurring in ML$'$. We will introduce in the model a set of closed terms $[\![\, \mathrm{Term} \,]\!]$ using essentially the same term and type constructors and destructors as in the type theory. α-equivalent terms (as well sets, context, judgements) will always be identified. However, we will, when introducing the models of the universes below occasionally throw away arguments needed for type checking purposes only (this is not really necessary, but simplifies matters).

One introduces as well a deterministic reduction relation $\longrightarrow$ on $[\![\, \mathrm{Term} \,]\!]$ corresponding to weak head reductions following the equality rules of the type theory. $\longrightarrow^*$ is the reflexive and transitive closure of $\longrightarrow$.

Definition 2.5. *(Environments)*

(a) An environment relative to a sequence of variables $x_1, \ldots, x_n$ is an expression $[x_1 = s_1, \ldots, x_n = s_n]$, where x_i are distinct variables and $s_i \in [\![\, \mathrm{Term} \,]\!]$.

(b) $\emptyset$ stands for the empty environment $[]$.

(c) $[x_1 = s_1, \ldots, x_n = s_n], y = t := [x_1 = s_1, \ldots, x_n = s_n, y = t]$*, where we assume that after some renaming of variables corresponding to* α*-conversion in the situation used,* y *is different from* x_i*.*

(d) $\mathrm{FV}([x_1 = s_1, \ldots, x_n = s_n]) := \{x_1, \ldots, x_n\}$*.*

(e) $[x_1 = s_1, \ldots, x_n = s_n](x_i) := s_i$*.*

(f) $\mathrm{Env}(x_1, \ldots, x_n) := \{[x_1 = s_1, \ldots, x_n = s_n] \mid s_1, \ldots, x_n \in [\![\, \mathrm{Term} \,]\!]\}$*.*

The interpretation $[\![\, t \,]\!]_\rho \in [\![\, \mathrm{Term} \,]\!]$ for environments ρ and terms $t \in \mathrm{Term}$ s.t. $\mathrm{FV}(t) \subseteq \mathrm{FV}(\rho)$ is the result of omitting in t the arguments of constructors/destructors to be thrown away (as indicated when introducing the constants) and substituting free variables x in t by $\rho(x)$.

We write $[\![\, t \,]\!]$ for $[\![\, t \,]\!]_\emptyset$.

Then one defines the interpretation $[\![\, A \,]\!]_\rho$ of sets A relative to environments ρ s.t. $\mathrm{FV}(A) \subseteq \mathrm{FV}(\rho)$ as a set of pairs $\langle s, t\rangle$ of closed terms. The intended meaning for $\langle s, t\rangle \in [\![\, A \,]\!]_\rho$ is that s and t are equal elements of the interpretation of A We write $[\![\, A \,]\!]$ for $[\![\, A \,]\!]_\emptyset$.

We define some basic operations on sets $X \subseteq [\![\, \mathrm{Term} \,]\!]^2$:

Definition 2.6. *Let* $X \subseteq [\![\, \mathrm{Term} \,]\!]^2$*.*

(a) The closure of X *under reductions is defined as*

$$\mathrm{Clos}_{\rightarrow}(X) := \{\langle s, t\rangle \mid \exists \langle s', t'\rangle \in X \mid s \longrightarrow^* s' \wedge t \longrightarrow^* t'\} \ .$$

(b) $\mathrm{Flat}(X) := \{a \in [\![\, \mathrm{Term} \,]\!] \mid \exists b \in [\![\, \mathrm{Term} \,]\!].\langle a, b\rangle \in X \vee \langle b, a\rangle \in X\}$*.*

(c) $\mathrm{PER}(X)$ *iff* X *is a partial equivalence relation (i.e. symmetric and transitive) and* X *is closed downwards under* $\longrightarrow$*, i.e. if* $s \longrightarrow^* s'$*,* $t \longrightarrow^* t'$ *and* $\langle s', t'\rangle \in X$*, then* $\langle s, t\rangle \in X$*.*

$\mathrm{Flat}([\![\, A \,]\!]_\rho)$ is the set of elements of the interpretation of A, and $\langle a, b\rangle \in [\![\, A \,]\!]_\rho$ means that a and b are equal elements of $[\![\, A \,]\!]_\rho$. This is meaningful if $\mathrm{PER}([\![\, A \,]\!]_\rho)$ holds, which we will show for all derivable sets A as expressed by the correctness statement below.

In the models for the type theories considered in this article, we can as well assign to every set A with $\mathrm{FV}(A) \subseteq \{x_1, \ldots, x_n\}$ an ordinal $\mathrm{o}(A) \geq 3$ s.t. for all $\rho \in \mathrm{Env}(x_1, \ldots, x_n)$ we have $[\![\, A \,]\!]_\rho \in \mathrm{L}'_{\mathrm{o}(A)}$, independent of ρ. Note that $[\![\, \mathrm{Term} \,]\!]$ is a Δ_0-definable subset of ω and therefore $[\![\, \mathrm{Term} \,]\!] \in \mathrm{L}'_3$.

The definition of $[\![\, A \,]\!]_\rho$ and $\mathrm{o}(A)$ for A formed by one of the standard set constructors is as follows:

- $[\![\, \mathrm{N} \,]\!]_\rho := [\![\, \mathrm{N} \,]\!] := \mathrm{Clos}_{\rightarrow}(\{\langle \mathrm{S}^n(0), \mathrm{S}^n(0)\rangle \mid n \in \omega\})$.
  $\mathrm{o}(\mathrm{N}) := 3$ (this is probably not the minimal ordinal one can associate with it, similarly for the later ordinals, but this bound suffices for our proof).

- $[\![\Sigma x : A.B]\!]_\rho := [\![\Sigma]\!]([\![A]\!]_\rho, \lambda s \in \mathrm{Flat}([\![A]\!]_\rho).[\![B]\!]_{\rho,x=s})$,
where, if $A \subseteq [\![\mathrm{Term}]\!]^2$, and for $a \in \mathrm{Flat}(A)$ we have $B(a) \subseteq [\![\mathrm{Term}]\!]^2$, then

$$\begin{aligned} [\![\Sigma]\!](A, B) := \mathrm{Clos}_{\rightarrow}(\{ \langle \mathrm{p}(a, b), \mathrm{p}(a', b') \rangle \\ \mid \langle a, a' \rangle \in A \wedge (B(a) = B(a')) \wedge \langle b, b' \rangle \in B(a)\}) \ . \end{aligned}$$

(Here $\lambda x \in A.s(x) := \{\langle x, s(x)\rangle \mid x \in A\}$).

$\mathrm{o}(\Sigma x : A.B) := \max\{\mathrm{o}(A), \mathrm{o}(B)\} + 2$.

- Similarly one defines the semantics and the ordinals of $\Pi x : A.B$, $\mathrm{W}x : a.B$, $A + B$, $\mathrm{Id}(A, a, b)$, $(x : A) \rightarrow B$, $(x : A) \times B$, making use of suitable definitions of $[\![\Pi]\!]$, $[\![\mathrm{W}]\!]$, $[\![+]\!]$, $[\![\mathrm{Id}]\!]$, $[\![\rightarrow]\!]$ and $[\![\times]\!]$, and incrementing the ordinals as for Σ by 2 (except for W, see below).

  We introduce here as well the following notations to be used later:

  - $(x \in A)\, [\![\rightarrow]\!]\, B := [\![\rightarrow]\!](A, \lambda x \in \mathrm{Flat}(A).B)$,
  - $A\, [\![\rightarrow]\!]\, B := (x \in A)\, [\![\rightarrow]\!]\, B$ where x is fresh,
  - $(x \in A)\, [\![\times]\!]\, B := [\![\times]\!](A, \lambda x \in \mathrm{Flat}(A).B)$,
  - $A\, [\![\times]\!]\, B := (x \in A)\, [\![\times]\!]\, B$ where x is fresh.
  - We can omit brackets in the above, using for $[\![\rightarrow]\!]$, $[\![\times]\!]$ the same conventions as we have for $\rightarrow$, $\times$.

  Only in case of W, one defines $\mathrm{o}(\mathrm{W}x : A.B) := (\max\{\mathrm{o}(A), \mathrm{o}(B)\})^+ + 1$, where α^+ is the next admissible ordinal above α. The reason for using $\alpha^+ + 1$ in case of $[\![\mathrm{W}]\!]$ is that $[\![\mathrm{W}]\!](A, B)$ is defined by iterating an operator, which makes the one step closure under the formation of new trees with subtrees being the previously defined trees up to α^+.

We interpret contexts as sets of environments as follows:

$$\begin{aligned} &[\![x_1 : A_1, \ldots, x_n : A_n]\!] := \\ &\{\langle [x_1 = s_1, \ldots, x_n = s_n], [x_1 = t_1, \ldots, x_n = t_n]\rangle \mid \\ &\qquad \langle s_1, t_1 \rangle \in [\![A_1]\!]_\emptyset \wedge \\ &\qquad \langle s_2, t_2 \rangle \in [\![A_2]\!]_{[x_1 = s_1]} \wedge \cdots \wedge \\ &\qquad \langle s_n, t_n \rangle \in [\![A_n]\!]_{[x_1 = s_1, \ldots, x_{n-1} = s_{n-1}]}\} \end{aligned}$$

Definition 2.7. *The correctness of a judgement is defined as follows by induction on the length:*

(a) $\mathrm{Correct}(\emptyset \Rightarrow \mathrm{Context})$ *(where $\emptyset$ is the empty context) is the true formula.*

(b)

$$\begin{aligned} &\mathrm{Correct}(\Gamma \Rightarrow A = B : \mathrm{Set}) \\ &:= \mathrm{Correct}(\Gamma \Rightarrow \mathrm{Context}) \wedge \forall \langle \rho, \rho' \rangle \in [\![\Gamma]\!].\mathrm{PER}([\![A]\!]_\rho) \wedge [\![A]\!]_\rho = [\![B]\!]_{\rho'} \ . \end{aligned}$$

(c)

$$\begin{aligned}&\mathrm{Correct}(\Gamma, y : A \Rightarrow \mathrm{Context}) := \mathrm{Correct}(\Gamma \Rightarrow A : \mathrm{Set})\\ &:= \mathrm{Correct}(\Gamma \Rightarrow A = A : \mathrm{Set})\ ,\end{aligned}$$

as just defined.

(d)

$$\begin{aligned}&\mathrm{Correct}(\Gamma \Rightarrow a = b : A)\\ &:= \mathrm{Correct}(\Gamma \Rightarrow A : \mathrm{Set}) \wedge \forall \langle \rho, \rho' \rangle \in [\![\, \Gamma \,]\!] . \langle [\![\, a \,]\!]_{\rho}, [\![\, a \,]\!]_{\rho'} \rangle \in [\![\, A \,]\!]_{\rho}\ .\end{aligned}$$

(e) $\mathrm{Correct}(\Gamma \Rightarrow a : A) := \mathrm{Correct}(\Gamma \Rightarrow a = a : A)$ *as just defined.*

The correctness statement to be shown by induction on the derivations in ML' is, that for Meta-every judgement $\Gamma \Rightarrow \theta$ we have

$$(\mathrm{ML}' \vdash \Gamma \Rightarrow \theta) \rightarrow \mathrm{KP}' \vdash \mathrm{Correct}(\Gamma \Rightarrow \theta)\ .$$

3 Universes and Recursively Inaccessibles

3.1 Rules for Universes

A universe (à la Tarski) in MLTT (Martin-Löf type theory) is given by a set $\mathrm{U} : \mathrm{Set}$ of codes for sets together with a decoding function $\mathrm{T} : \mathrm{U} \rightarrow \mathrm{Set}$, which without the logical framework is given by the rule:

$$\frac{u : \mathrm{U}}{\mathrm{T}(u) : \mathrm{Set}}$$

All universes in the following will be closed under the basic set constructions (i.e. N_n, N, Σ, Π, W, +, Id). This means that we have in case of N and Σ the rules:

$$\widehat{\mathrm{N}} : \mathrm{U} \qquad\qquad \mathrm{T}(\widehat{\mathrm{N}}) = \mathrm{N}$$

$$\frac{a : \mathrm{U} \qquad b : \mathrm{T}(a) \rightarrow \mathrm{U}}{\widehat{\Sigma}(a, b) : \mathrm{U}} \qquad \mathrm{T}(\widehat{\Sigma}(a, b)) = \Sigma x : \mathrm{T}(a).\mathrm{T}(b(x))$$

(The codes for N_n, Π, W, +, Id will be called $\widehat{\mathrm{N}}_n$, $\widehat{\Pi}$, $\widehat{\mathrm{W}}$, $\widehat{+}$, $\widehat{\mathrm{Id}}$ respectively.)

In particular we consider the theory $\mathrm{ML_WU}$ consisting of the small logical framework, the basic set constructions and one universe (which is by our convention closed under the standard set constructors including the W-set, but not under U itself; the subscript W stands for the closure under the W-set).

We will not add any elimination rules for universes. In case of the theory $\mathrm{ML_WU}$ and of the super universe (as introduced in Sect. 4), they could be added without any problems, but won't add any strength to them. In case of the Mahlo universe (Sect. 5) and stronger universes, such elimination rules result in inconsistencies (see Theorem 6.1 in [20]).

Type theories with several universes. We will already here look at how to extend this approach to type theories having several universes. In this case, the codes $\widehat{\mathrm{N}}, \widehat{\Sigma}$ etc. have special names and additional parameters so that we can distinguish them from each other. These parameters will always be omitted when moving to $[\![\,\mathrm{Term}\,]\!]$. E.g. if we have universes V and $\mathrm{U}_{a,b}$, we have in Term codes $\widehat{\mathrm{N}}_{\mathrm{V}}$ and $\widehat{\mathrm{N}}_{\mathrm{U},a,b}$, whereas in $[\![\,\mathrm{Term}\,]\!]$ we only have a code $\widehat{\mathrm{N}}$ and have $[\![\,\widehat{\mathrm{N}}_{\mathrm{V}}\,]\!]_\rho = [\![\,\widehat{\mathrm{N}}_{\mathrm{U},a,b}\,]\!]_\rho = \widehat{\mathrm{N}}$.

By a *recursive subuniverse* $(\mathrm{U}, \widehat{\mathrm{T}}_{\mathrm{U}})$ of a universe $(\mathrm{V}, \mathrm{T}_{\mathrm{V}})$ we mean that we have $\mathrm{U} : \mathrm{Set}$, but, instead of having $\mathrm{T}_{\mathrm{U}} : \mathrm{U} \to \mathrm{Set}$ (or the corresponding formulation above avoiding the full logical framework), have the rule formulating the existence of a function $\widehat{\mathrm{T}}_{\mathrm{U}} : \mathrm{U} \to \mathrm{V}$, which recursively decodes codes by their corresponding codes in the other set. If we for $a : \mathrm{U}$ define $\mathrm{T}_{\mathrm{U}}(a) := \mathrm{T}_{\mathrm{V}}(\widehat{\mathrm{T}}_{\mathrm{U}}(a)) :$ Set, we have therefore the rules:

$$\widehat{\mathrm{N}}_{\mathrm{U}} : \mathrm{U} \qquad\qquad \widehat{\mathrm{T}}_{\mathrm{U}}(\widehat{\mathrm{N}}_{\mathrm{U}}) = \widehat{\mathrm{N}}_{\mathrm{V}}$$

$$\frac{a : \mathrm{U} \qquad b : \mathrm{T}_{\mathrm{U}}(a) \to \mathrm{U}}{\widehat{\Sigma}_{\mathrm{U}}(a, b) : \mathrm{U}} \quad \widehat{\mathrm{T}}_{\mathrm{U}}(\widehat{\Sigma}_{\mathrm{U}}(a, b)) = \widehat{\Sigma}_{\mathrm{V}}(\widehat{\mathrm{T}}_{\mathrm{U}}(a), \widehat{\mathrm{T}}_{\mathrm{U}} \circ b)$$

similarly for the other basic set constructions.

Sometimes it will not be possible to develop the subuniverses introduced in this article as recursive subuniverses. The problem occurs as soon as we have infinite chains of universes above one universe. Consider for instance a sequence of universes $(\mathrm{U}_\alpha)_{\alpha \le \beta}$. There is no problem, if we have a finite sequence of such universes (i.e. $\beta < \omega$): Then we can consider for $\alpha < \beta$ the universe U_α as a recursive subuniverse of $\mathrm{U}_{\alpha+1}$, whereas the universe U_β itself is defined as a standard universe. The set corresponding to $a : \mathrm{U}_0$ is then $\mathrm{T}_\beta(\widehat{\mathrm{T}}_{\beta-1}(\cdots(\widehat{\mathrm{T}}_0(a))\cdots))$. If $\beta = \omega$, we would with this approach end up with a nonterminating chain

$$\mathrm{U}_0 \xrightarrow{\widehat{\mathrm{T}}_0} \mathrm{U}_1 \xrightarrow{\widehat{\mathrm{T}}_1} \mathrm{U}_2 \xrightarrow{\widehat{\mathrm{T}}_2} \cdots$$

and we cannot determine a set corresponding to $x : \mathrm{U}_0$.

The way around this is to follow Per Martin-Löf's approach, by forming *inductive subuniverses*. An inductive subuniverse U of a universe $(\mathrm{V}, \mathrm{T}_{\mathrm{V}})$ is given by

- a set U,
- a recursively defined decoding function $\mathrm{T}_{\mathrm{U}} : \mathrm{U} \to \mathrm{Set}$ and
- a *constructor* $\widehat{\mathrm{T}}_{\mathrm{U}} : \mathrm{U} \to \mathrm{V}$.

So $\widehat{\mathrm{T}}_{\mathrm{U}} : \mathrm{U} \to \mathrm{V}$ forms an additional introduction rule for V.

If we avoid the full logical framework, we get the following rules for an inductive subuniverse:

$$\mathrm{U} : \mathrm{Set} \qquad \frac{a : \mathrm{U}}{\mathrm{T}_{\mathrm{U}}(a) : \mathrm{Set}}$$

$$\frac{a : \mathrm{U}}{\widehat{\mathrm{T}}_{\mathrm{U}}(a) : \mathrm{V}} \qquad \mathrm{T}_{\mathrm{V}}(\widehat{\mathrm{T}}_{\mathrm{U}}(a)) = \mathrm{T}_{\mathrm{U}}(a)$$

$$\widehat{\mathrm{N}}_{\mathrm{U}} : \mathrm{U} \qquad \mathrm{T}_{\mathrm{U}}(\widehat{\mathrm{N}}_{\mathrm{U}}) = \mathrm{N}$$

$$\frac{a : \mathrm{U} \qquad b : \mathrm{T}_{\mathrm{U}}(a) \to \mathrm{U}}{\widehat{\Sigma}_{\mathrm{U}}(a, b) : \mathrm{U}} \quad \mathrm{T}_{\mathrm{U}}(\widehat{\Sigma}_{\mathrm{U}}(a, b)) = \Sigma x : \mathrm{T}_{\mathrm{U}}(a).\mathrm{T}_{\mathrm{U}}(b(x))$$

Similarly for the other basic set constructions.

The reason why we prefer recursive subuniverses to inductive ones is that we obtain in inductive subuniverses many copies of the same element. In the simple situation, we have for instance two codes for N in V, namely $\widehat{\mathrm{N}}_{\mathrm{V}}$ and $\widehat{\mathrm{T}}_{\mathrm{V}}(\widehat{\mathrm{N}}_{\mathrm{U}})$. This is not very aesthetic (it doesn't cause many problems when actually making use of such universes). However, one should be aware that we anyway get often doubling of codes. For instance, if in case of the super universe to be introduced later in this article we form a recursive subuniverse $(\mathrm{U}, \widehat{\mathrm{T}}_{\mathrm{U}})$ of $\mathrm{V}, \mathrm{T}_{\mathrm{V}}$ containing the family $\langle \mathrm{N}_{\mathrm{V}}, (x)\widehat{\mathrm{N}}_{\mathrm{V}} \rangle$, then U will have infinitely many codes for N, namely $\widehat{\mathrm{N}}_{\mathrm{U}}$, $\widehat{\mathrm{a}}$ and $\widehat{\mathrm{b}}(n)$ for $n : \mathrm{N}$.

When modelling subuniverses, we will treat them as mere subsets of the interpretation of the original universe. $\widehat{\mathrm{T}}_{\mathrm{U}}$ will for both the recursive and inductive subuniverses be essentially the identity function, expressed by the reduction rule $\widehat{\mathrm{T}}_{\mathrm{U}}(a) \longrightarrow a$. This means that in the model we will identify codes coming from the same element in U, so for instance all the codes for the natural numbers in the above example will be equal in the model.

3.2 Modelling Universes

For modelling universes closed under W in extensions of KPω one needs the notion of a recursively inaccessible. A recursively inaccessible ordinal I is an admissible ordinal I s.t. for every $\beta \in \mathrm{I}$ there exists a $\gamma \in \mathrm{I}$ s.t. $\beta \in \gamma$ and γ is an admissible ordinal. If one refers to admissible sets rather than admissible ordinals, then an recursively inaccessible set is a set a s.t.

$$\mathrm{Ad}(a) \wedge \forall y \in a.\exists z \in a.y \in z \wedge \mathrm{Ad}(z) \ .$$

Let KPI$^+$ be the theory of KPω extended by the existence of one recursively inaccessible set and finitely many admissibles above it. So KPI$^+$ has constants $\mathrm{a}_{\mathrm{I},k}$ for Meta $k \in \omega$, axioms expressing that $\mathrm{a}_{\mathrm{I},0}$ is a recursively inaccessible set and for Meta $k \in \omega$ the axioms

$$\mathrm{Ad}(\mathrm{a}_{\mathrm{I},k}) \wedge \mathrm{a}_{\mathrm{I},k} \in \mathrm{a}_{\mathrm{I},k+1} \ .$$

Proof theoretically equivalent is the theory KPl plus the existence of one recursively inaccessible set.

A universe is a family of sets, and therefore it will be modelled by a family of terms $[\![\mathrm{Fam}]\!](\mathrm{set})$, as given by the following definition:

Definition 3.1. *(in the language of Kripke-Platek set theory).*

(a) *By* set *we mean the class of all sets.*

(b) $$\begin{aligned}[\![\mathrm{Fam}]\!](\mathrm{set}) := \{(A,B) \in \mathrm{set}^2 \mid & A \subseteq [\![\mathrm{Term}]\!]^2 \\ & \wedge B \text{ is a function} \\ & \wedge \mathrm{dom}(B) = \mathrm{Flat}(A) \\ & \wedge \forall x \in \mathrm{Flat}(A).B(x) \subseteq [\![\mathrm{Term}]\!]^2 \\ & \wedge \forall \langle x, x' \rangle \in A.B(x) = B(x')\}.\end{aligned}$$

(c) *If* $(A,B) \in [\![\mathrm{Fam}]\!](\mathrm{set})$. *Then*

$$[\![\mathrm{Fam}]\!](A,B) := (x \in A)\, [\![\times]\!]\, (B(x)\, [\![\to]\!]\, A) \ ,$$

(which is

$$\begin{aligned}\mathrm{Clos}_{\to}(\{\langle c, c' \rangle \in [\![\mathrm{Term}]\!]^2 \mid & \langle \pi_0(c), \pi_0(c') \rangle \in A \\ & \wedge (B(\pi_0(c))\, [\![\to]\!]\, A) = (B(\pi_0(c'))\, [\![\to]\!]\, A) \\ & \wedge \forall \langle x, x' \rangle \in B(\pi_0(c)).\langle \pi_1(c)(x), \pi_1(c')(x') \rangle \in A\})\ ) \ .\end{aligned}$$

(d) *If* (A,B), $(A',B') \in [\![\mathrm{Fam}]\!](\mathrm{set})$. *Then*

$$(A,B) \le (A',B') :\Leftrightarrow A \subseteq A' \wedge B \subseteq B' \ .$$

Note that by (A,B), $(A',B') \in [\![\mathrm{Fam}]\!](\mathrm{set})$, $B \subseteq B'$ *is equivalent to*

$$B \restriction \mathrm{Flat}(A) = B' \restriction \mathrm{Flat}(A) \ .$$

(e) *If* $(U^\alpha, T^\alpha) \in [\![\mathrm{Fam}]\!](\mathrm{set})$ *for* $\alpha < \beta$ *s.t.*

$$\forall \alpha < \alpha' < \beta.(U^\alpha, T^\alpha) \le (U^{\alpha'}, T^{\alpha'}) \ .$$

Then we define

$$(U^{<\beta}, T^{<\beta}) := (\bigcup_{\alpha<\beta} U^\alpha, \bigcup_{\alpha<\beta} T^\alpha) \ .$$

Note that, if $x \in \mathrm{Flat}(U^\alpha)$, $\alpha < \beta$, *then*

$$T^{<\beta}(x) = T^\alpha(x) \ .$$

A model of a universe (U, T) can be obtained by defining sets $[\![\mathrm{U}]\!]$ and $[\![\mathrm{T}]\!]$ s.t.

$$([\![\mathrm{U}]\!], [\![\mathrm{T}]\!]) \in [\![\mathrm{Fam}]\!](\mathrm{set}) \ .$$

Then

$$[\![\mathrm{T}(s)]\!]_\rho := [\![\mathrm{T}]\!]([\![s]\!]_\rho) \ .$$

We will introduce first by induction on α sets

$$(\mathrm{U}^\alpha, \mathrm{T}^\alpha) \in [\![\, \mathrm{Fam} \,]\!](\mathrm{set})$$

s.t.

$$\alpha < \beta \to (\mathrm{U}^\alpha, \mathrm{T}^\alpha) < (\mathrm{U}^\beta, \mathrm{T}^\beta) \ \ .$$

then we define (assuming the existence of a recursively inaccessible ordinal I)

$$[\![\, \mathrm{U} \,]\!] = \mathrm{U}^{<\mathrm{I}} \ \ , \ \ [\![\, \mathrm{T} \,]\!] = \mathrm{U}^{<\mathrm{I}}$$

for the least recursively inaccessible ordinal I. (In case we close the universe under additional constructions I will be replaced by a recursively inaccessible ordinal, which has some additional closure properties.)

We will guarantee that

$$\mathrm{PER}(\mathrm{U}^\alpha) \wedge \forall x \in \mathrm{Flat}(\mathrm{U}^\alpha)\mathrm{PER}(\mathrm{T}^\alpha(x)) \ \ .$$

We will as well enforce

$$\mathrm{U}^\alpha \in \mathrm{L}'_{\alpha+1} \wedge \forall x \in \mathrm{Flat}(\mathrm{U}^\alpha)\mathrm{T}^\alpha(x) \in \mathrm{L}'_\alpha \ \ .$$

All approximations of interpretations of universes U^α introduced in this article will be closed under reductions, which means in the following:

- If $\langle b, b' \rangle \in \mathrm{U}^\alpha$, $c \longrightarrow^* b$, $c' \longrightarrow^* b'$ then $\langle c, c' \rangle \in \mathrm{U}^\alpha$.
- If $b \longrightarrow^* c \in \mathrm{U}^\alpha$, then $\mathrm{T}^\alpha(b) = \mathrm{T}^\alpha(c)$.

More formally the definition of U^α below has to be replaced by the definition of a set U'^α and function T'^α, and one defines then

$$\begin{array}{c} \mathrm{U}^\alpha := \mathrm{Closure}(\mathrm{U}'^\alpha) \ \ , \\ \mathrm{T}^\alpha(a) := \mathrm{T}'^\alpha(a') \text{ if } a \longrightarrow^* a' \in \mathrm{U}'^\alpha \ \ . \end{array}$$

We will however in the following not explicitly mention U'^α, T'^α in this article, assuming that the reader can adapt the model correspondingly.

U^α is the closure of $\mathrm{U}^{<\alpha}$ under one application of each basic set constructions of type theory, provided the results are in L'_α:

- $[\![\, \mathrm{N} \,]\!] \in \mathrm{L}'_\alpha \to \langle \widehat{\mathrm{N}}, \widehat{\mathrm{N}} \rangle \in \mathrm{U}^\alpha$, $\mathrm{T}^\alpha(\widehat{\mathrm{N}}) = [\![\, \mathrm{N} \,]\!]$.
- If $\langle \langle a, b \rangle, \langle a' b' \rangle \rangle \in [\![\, \mathrm{Fam} \,]\!](\mathrm{U}^{<\alpha}, \mathrm{T}^{<\alpha})$, and

  $$A := [\![\, \Sigma \,]\!](\mathrm{T}^{<\alpha}(a), \lambda x \in \mathrm{T}^{<\alpha}(a).\mathrm{T}^{<\alpha}(b(y))) \in \mathrm{L}'_\alpha \ \ ,$$

  then $\langle \widehat{\Sigma}(a, b), \widehat{\Sigma}(a', b') \rangle \in \mathrm{U}^\alpha$ and $\mathrm{T}^\alpha(\widehat{\Sigma}(a, b)) = A$.
- Similarly for the other standard constructions of type theory.

In order to get a model of universe closed under some other principles, one will later add as well to U^α elements corresponding to those sets. In this article, these closure principles will always be the closure under certain universes.

We show that $(\mathrm{U}^{<\mathrm{I}}, \mathrm{T}^{<\mathrm{I}})$ is a model of the standard universe as follows (The proof shows in general that, if we close U^α under additional constructions in such a way that $\mathrm{U}^\alpha \in \mathrm{L}'_{\alpha+1}$ and for $x \in \mathrm{Flat}(\mathrm{U}^\alpha)$, $\mathrm{T}^\alpha(x) \in \mathrm{L}'_\alpha$, and if κ is a recursively inaccessible, then $(\mathrm{U}^{<\kappa}, \mathrm{T}^{<\kappa})$ is closed under the basic set constructions):

First one observes that if $A \subseteq [\![\,\mathrm{Term}\,]\!]^2$, $A \in \mathrm{L}'_\alpha$, for $a \in \mathrm{Flat}(A)$, $B(x) \in [\![\,\mathrm{Term}\,]\!]^2$ and $B(x) \in \mathrm{L}'_\alpha$, then $[\![\,\Sigma\,]\!](A, B)$, $[\![\,\Pi\,]\!](A, B) \in \mathrm{L}'_{\alpha+2}$ and $[\![\,\mathrm{W}\,]\!](A, B) \in \mathrm{L}'_{\alpha^++1}$. Similar properties hold for $[\![\,+\,]\!]$ and $[\![\,\mathrm{Id}\,]\!]$ (making use of $\alpha + 2$).

Now, if I is a recursively inaccessible ordinal, then $(\mathrm{U}^{<\mathrm{I}}, \mathrm{T}^{<\mathrm{I}})$ is closed under $\widehat{\mathrm{N}}_n$, $\widehat{\mathrm{N}}$, Σ, Π, $\widehat{\mathrm{W}}$, $\widehat{+}$, $\widehat{\mathrm{Id}}$: We consider only the most difficult case $\widehat{\mathrm{W}}$: First we show a small Lemma, which will be useful in general and will hold for all universe constructions defined in this article:

Lemma 3.2. *Let κ be admissible, and*

$$\langle\langle a, b\rangle, \langle a', b'\rangle\rangle \in [\![\,\mathrm{Fam}\,]\!](\mathrm{U}^{<\kappa}, \mathrm{T}^{<\kappa}) \enspace .$$

then there exists an $\alpha < \kappa$ s.t.

$$\langle\langle a, b\rangle, \langle a', b'\rangle\rangle \in [\![\,\mathrm{Fam}\,]\!](\mathrm{U}^{<\alpha}, \mathrm{T}^{<\alpha}) \enspace .$$

Proof of the Lemma: We have $\langle a, a'\rangle \in \mathrm{U}^{<\kappa}$, therefore there exists $\alpha_0 < \kappa$ s.t. $\langle a, a'\rangle \in \mathrm{U}^{<\alpha_0}$. Furthermore, $\mathrm{T}^{<\alpha_0}(a) \in \mathrm{L}'_{\alpha_0}$, and for all $\langle x, x'\rangle \in \mathrm{T}^{<\alpha_0}(a)$ there exists $\alpha_1 < \kappa$ s.t. $\langle b(x), b'(x')\rangle \in \mathrm{U}^{<\alpha_1}$. By κ admissible we can find a uniform α_1 for this. Now let $\alpha := \max\{\alpha_0, \alpha_1\}$.

We continue with the main proof. Assume $\langle\langle a, b\rangle, \langle a', b'\rangle\rangle \in [\![\,\mathrm{Fam}\,]\!](\mathrm{U}^{<\mathrm{I}}, \mathrm{T}^{<\mathrm{I}})$. Then by Lemma 3.2 we find an α s.t. $\langle\langle a, b\rangle, \langle a', b'\rangle\rangle \in [\![\,\mathrm{Fam}\,]\!](\mathrm{U}^{<\alpha}, \mathrm{T}^{<\alpha})$. We have $\mathrm{T}^{<\mathrm{I}}(a) = \mathrm{T}^{<\beta}(a) \in \mathrm{L}'_\beta$, $\mathrm{T}^{<\mathrm{I}}(b(x)) = \mathrm{T}^{<\beta}(b(x)) \in \mathrm{L}'_\beta$, therefore $[\![\,\mathrm{W}\,]\!](\mathrm{T}^{<\mathrm{I}}(a), \lambda x \in \mathrm{T}^{<\mathrm{I}}(a).\mathrm{T}^{<\mathrm{I}}(b(x))) \in \mathrm{L}'_{\beta^++1}$. Therefore $\langle\widehat{\mathrm{W}}(a, b), \widehat{\mathrm{W}}(a', b')\rangle \in \mathrm{U}^{\beta^++2} \subseteq \mathrm{U}^{<\mathrm{I}}$.

One obtains a model of the type theory $\mathrm{ML_WU}$ in KPI^+ as follows: We have modelled all the ingredients already, and need only to make sure that the constructions can be carried out in KPI^+. Let κ_n be the nth admissible above I. First, one observes, that $[\![\,\mathrm{U}\,]\!] \in \mathrm{L}'_{\mathrm{I}+2}$, for $x \in \mathrm{Flat}([\![\,\mathrm{U}\,]\!])$ we have $[\![\,\mathrm{T_U}\,]\!](x) \in \mathrm{L}'_\mathrm{I}$, and for $\langle x, x'\rangle \in [\![\,\mathrm{U}\,]\!]$ we have $[\![\,\mathrm{T_U}\,]\!](x) = [\![\,\mathrm{T_U}\,]\!](x')$. Therefore one defines $\mathrm{o}(\mathrm{U}) := \mathrm{I} + 2$, $\mathrm{o}(\mathrm{T_U}(s)) := \mathrm{I}$. For all other set constructions we have seen that if the maximum of the ordinals of the subterms is α, then the ordinal of the term itself is $\alpha + 2$ or $\alpha^+ + 1$. Therefore we obtain that for any set term A there exists an n s.t. $\mathrm{o}(A) \leq \kappa_n$ and therefore for all ρ $[\![\,A\,]\!]_\rho \in \mathrm{L}'_{\kappa_n}$. One sees now that we can interpret all sets in KPI^+, and can show that this model is sound with respect to the above mentioned correctness conditions. So we have given a sketch of the following theorem of which a detailed proof can be found in [25]; see as well [27]; Griffor and Rathjen have shown related results in [14]:

Theorem 3.3.

(a) $\mathrm{ML_W U}$ *can be modelled in* KPI^+.

(b) $|\mathrm{ML_W U}| \leq |\mathrm{KPI}^+|$.

(c) *The previous statements hold as well if we replace intensional by extensional equality.*

In the other direction, we have developed in [29] (see as well our PhD thesis [25], the overview articles [28] and [31]; Rathjen and Griffor have obtained related results in [14]) a well-ordering proof which shows the other direction, namely:

Theorem 3.4. *(a)* $|\mathrm{ML_W U}| = |\mathrm{KPI}^+|$

(b) *The previous statements hold as well if we replace intensional by extensional equality.*

4 Palmgren's Super Universe and Recursively Hyperinaccessibles

4.1 Definition of the Super Universe

The first substantial step beyond the type theory $\mathrm{ML_W U}$ was the introduction of a hierarchy of finitely many universes by Per Martin-Löf. We won't cover this extension in this article, and will instead move to the next step, namely the super universe, as introduced by Erik Palmgren. He introduced a super universe operator, which forms a universe containing a family of sets. The super universe is defined as a universe closed under this operator.

In order to introduce this in detail, let us first introduce some notions:

Definition 4.1. *(Refers to the logical framework).*

(a) *The type of families of sets is defined (using the logical framework) as* $\mathrm{Fam}(\mathrm{Set}) := (X : \mathrm{Set}) \times (X \to \mathrm{Set})$.

(b) *Let* $\langle \mathrm{U}, \mathrm{T}\rangle \in \mathrm{Fam}(\mathrm{Set})$. *The family of sets in* $\langle \mathrm{U}, \mathrm{T}\rangle$ *is the set of families of sets in* U*, indexed over elements of* U*:*
$\mathrm{Fam}(\mathrm{U}, \mathrm{T}) := (x : \mathrm{U}) \times (\mathrm{T}(x)) \to \mathrm{U})$.

(c) *Let* $\langle \mathrm{U}, \mathrm{T}\rangle \in \mathrm{Fam}(\mathrm{Set})$. *The lifting of an element* $a : \mathrm{Fam}(\mathrm{U}, \mathrm{T})$ *to an element of* $\mathrm{Fam}(\mathrm{Set})$ *is defined as* $\mathrm{T}[a] := \langle \mathrm{T}(\pi_0(a)), (x)\mathrm{T}(\pi_1(a)(x))\rangle : \mathrm{Fam}(\mathrm{Set})$.

Palmgren, first introduced the super universe operator

$$\mathrm{SU} : \mathrm{Fam}(\mathrm{Set}) \to \mathrm{Fam}(\mathrm{Set}) \ .$$

$\mathrm{SU}(A)$ is a universe $\langle U', T'\rangle$ closed under the standard set constructions and containing A. This means that if $A = \langle A', B'\rangle$, then U' contains

- a code $\widehat{\mathrm{a}}$ for A', i.e. $T'(\widehat{\mathrm{a}}) = A'$,
- and for $a : A'$ contains a code $\widehat{\mathrm{b}}(a)$ for $B'(a)$, i.e. $T'(\widehat{\mathrm{b}}(a)) = B'(a)$.

The next step was to introduce the super universe as a universe $(\mathrm{V}, \mathrm{T}_{\mathrm{V}})$, which is as usual closed under the standard set constructions and which reflects the super universe operator: if $a : \mathrm{Fam}(\mathrm{V}, \mathrm{T}_{\mathrm{V}})$, then V contains codes $\langle \widehat{\mathrm{U}}_a, (x)\widehat{\mathrm{T}}_a(x)\rangle$ for $\mathrm{SU}(\mathrm{T}_{\mathrm{V}}[a])$, which means that

$$\mathrm{T}_{\mathrm{V}}[\langle \widehat{\mathrm{U}}_a, (x)\widehat{\mathrm{T}}_a(x)\rangle] = \mathrm{SU}(\mathrm{T}_{\mathrm{V}}[a]) \ .$$

We will in the following present a version, in which $(\mathrm{T}(\widehat{\mathrm{U}}_a), \widehat{\mathrm{T}}_a)$ is a recursive subuniverse of $(\mathrm{V}, \mathrm{T}_{\mathrm{V}})$. Furthermore, we will define the universe operator only restricted to elements of $\mathrm{Fam}(\mathrm{V}, \mathrm{T}_{\mathrm{V}})$ (which will be written as $\mathrm{U}_{a,b}$ for $\langle a, b\rangle : \mathrm{Fam}(\mathrm{V}, \mathrm{T}_{\mathrm{V}})$). We uncurry SU and obtain the following rules:

$$\mathrm{V} : \mathrm{Set} \qquad \frac{a : \mathrm{V}}{\mathrm{T}_{\mathrm{V}}(a) : \mathrm{Set}}$$

$$\frac{a : \mathrm{V} \qquad b : \mathrm{T}_{\mathrm{V}}(a) \to \mathrm{V}}{\mathrm{U}_{a,b} : \mathrm{Set}}$$

$$\frac{a : \mathrm{V} \qquad b : \mathrm{T}_{\mathrm{V}}(a) \to \mathrm{V}}{\widehat{\mathrm{U}}_{a,b} : \mathrm{V}} \qquad \mathrm{T}_{\mathrm{V}}(\widehat{\mathrm{U}}_{a,b}) = \mathrm{U}_{a,b}$$

$$\frac{a : \mathrm{V} \qquad b : \mathrm{T}_{\mathrm{V}}(a) \to \mathrm{V} \qquad c : \mathrm{U}_{ab}}{\widehat{\mathrm{T}}_{\mathrm{U},a,b}(c) : \mathrm{V}}$$

Let $\mathrm{T}_{\mathrm{U},a,b}(x) := \mathrm{T}_{\mathrm{V}}(\widehat{\mathrm{T}}_{\mathrm{U},a,b}(x)) : \mathrm{Set}$.

$(\mathrm{V}, \mathrm{T}_{\mathrm{V}})$ is a universe, which means it is closed under the basic set constructions. The codes are written as $\widehat{\mathrm{N}}_{\mathrm{V}}$, $\widehat{\Sigma}_{\mathrm{V}}$, etc.

One demands for $a : \mathrm{V}$, $b : \mathrm{T}_{\mathrm{V}}(a) \to \mathrm{V}$ that $(\mathrm{U}_{a,b}, \widehat{\mathrm{T}}_{\mathrm{U},a,b})$ is a recursive subuniverse of $(\mathrm{V}, \mathrm{T}_{\mathrm{V}})$, which means as indicated in Sect. 3 that $\mathrm{U}_{a,b}$ is closed under constructors for forming codes for the standard set constructions like $\widehat{\mathrm{N}}_{\mathrm{U},a,b}$, $\widehat{\Sigma}_{\mathrm{U},a,b}$, etc., and that $\widehat{\mathrm{T}}_{\mathrm{U},a,b}$ decodes them as the corresponding codes in V, e.g. $\widehat{\mathrm{T}}_{\mathrm{U},a,b}(\widehat{\mathrm{N}}_{\mathrm{U},a,b}) = \widehat{\mathrm{N}}_{\mathrm{V}}$.

Furthermore, $\mathrm{U}_{a,b}$ contains codes for a, b, which means that one has the following rules:

$$\widehat{\mathrm{a}}_{a,b} : \mathrm{U}_{a,b} \qquad \widehat{\mathrm{T}}_{\mathrm{U},a,b}(\widehat{\mathrm{a}}_{a,b}) = a$$

$$\frac{c : \mathrm{T}_{\mathrm{U},a,b}(a)}{\widehat{\mathrm{b}}_{a,b}(c) : \mathrm{U}_{a,b}} \quad \widehat{\mathrm{T}}_{\mathrm{U},a,b}(\widehat{\mathrm{b}}_{a,b}(c)) = b(c)$$

Let ML_{W} + (Superuniv) be the type theory consisting of the rules of the small logical framework, the basic set constructions, and the rules for V, T_{V}, U_{ab}, $\widehat{\mathrm{U}}_{ab}$, $\widehat{\mathrm{T}}_{\mathrm{U},a,b}$ as above.

In this type theory, one can define using the elimination rules for N a hierarchy of universes $\widehat{\mathrm{U}}'_n : \mathrm{V}$ by

$$\widehat{\mathrm{U}}'_0 = \widehat{\mathrm{U}}_{\widehat{\mathrm{N}}_0,(x)\widehat{\mathrm{N}}_0} \quad \widehat{\mathrm{U}}'_{\mathrm{T}_\mathrm{V}(n)} = \widehat{\mathrm{U}}_{\widehat{\mathrm{N}}_1,(n)\widehat{\mathrm{U}}'_n}$$

but as well a universe containing this hierarchy of universes, namely $\mathrm{U}_{\widehat{\mathrm{N}},(n)\widehat{\mathrm{U}}'_n}$.

This is the first step towards forming a transfinite hierarchy of universes: for any well-founded relation (with well-foundedness provable in this type theory) $R, \prec$ we can define a hierarchy of universes $\widehat{\mathrm{U}}''_\alpha$ for $\alpha \in R$ s.t. $\widehat{\mathrm{U}}''_\alpha$ contains codes for $\widehat{\mathrm{U}}''_\beta$ for $\beta \prec \alpha$.

4.2 A Model of the Super Universe

$\mathrm{ML_W}$ + (Superuniv) has strength $(\mathrm{KP} + (\mathrm{hyper} - \mathrm{inacc}))^+$, where $(\mathrm{KP} + (\mathrm{hyper} - \mathrm{inacc}))^+$ is Kripke-Platek set theory plus the existence of one recursively hyper-inaccessible set plus finitely many admissibles above it, which is equivalent to KPl plus the existence of one recursively hyper-inaccessible set.

Here a recursively hyper-inaccessible set is an admissible set a s.t. for all $x \in a$ there exists a $y \in a$ s.t. $x \in y$ and y is recursively inaccessible. A recursively hyper-inaccessible ordinal is an admissible ordinal HI s.t.

$$\forall \alpha \in \mathrm{HI}. \exists \beta \in \mathrm{HI}. \alpha \in \beta \wedge \mathrm{Inacc}(\beta) \enspace ,$$

where $\mathrm{Inacc}(\beta)$ means that β is a recursively inaccessible ordinal.

The precise formulation of $(\mathrm{KP} + (\mathrm{hyper} - \mathrm{inacc}))^+$ is similar to KPI^+, except that one replaces "inaccessible" by "hyper-inaccessible".

An upper bound for the strength of $\mathrm{ML_W}$ + (Superuniv) can be obtained by modelling it in $(\mathrm{KP} + (\mathrm{hyper} - \mathrm{inacc}))^+$. We will give in the following a sketch of the model, full details will be presented in a future paper.

The basic construction is outlined in Sect. 2. As pointed out there, in $[\![\,\mathrm{Term}\,]\!]$ we forget about the constants used for type checking only, so we have for instance only a constant $\widehat{\mathrm{N}}$ and interpret

$$[\![\,\widehat{\mathrm{N}}_\mathrm{V}\,]\!]_\rho = [\![\,\widehat{\mathrm{N}}_{\mathrm{U},a,b}\,]\!]_\rho = \widehat{\mathrm{N}} \enspace ,$$

similarly for $\widehat{\Sigma}$, $\widehat{\mathrm{a}}$, $\widehat{\mathrm{b}}$. $\widehat{\mathrm{U}}$, $\widehat{\mathrm{T}}_\mathrm{U}$ will depend on a, b, and we will add the reduction rules for $\widehat{\mathrm{T}}_\mathrm{V}$, expressed by the above, i.e.

$$\begin{aligned}
&\widehat{\mathrm{T}}_{\mathrm{U},a,b}(\widehat{\mathrm{N}}) \longrightarrow \widehat{\mathrm{N}} \enspace ,\\
&\widehat{\mathrm{T}}_{\mathrm{U},a,b}(\widehat{\Sigma}(c,d)) \longrightarrow \widehat{\Sigma}(c,d) \enspace ,\\
&\widehat{\mathrm{T}}_{\mathrm{U},a,b}(\widehat{\mathrm{a}}) \longrightarrow a \enspace ,\\
&\widehat{\mathrm{T}}_{\mathrm{U},a,b}(\widehat{\mathrm{b}}(c)) \longrightarrow b(c) \enspace .
\end{aligned}$$

As for standard universes, the super universe is interpreted, by defining first by induction over α sets

$$(\mathrm{V}^\alpha, \mathrm{T}^\alpha) \in [\![\,\mathrm{Fam}\,]\!](\mathrm{set})$$

s.t.

$$\alpha < \beta \to (\mathrm{U}_\alpha, \mathrm{T}_\alpha) < (\mathrm{U}_\beta, \mathrm{T}_\beta) \ .$$

As for all models of universes in this article, we add to V^α the codes for standard set constructors applied to arguments, if the elements of the universe they refer to are already in $\mathrm{V}^{<\alpha}$ and the sets added are in L'_α. Additionally, whenever there exists a β s.t. $\beta+1 < \alpha$ and $\mathrm{V}^{<\beta}$ is closed under all universe constructions, and s.t.

$$\langle\langle a, b\rangle, \langle a', b'\rangle\rangle \in [\![\,\mathrm{Fam}\,]\!](\mathrm{V}^{<\beta}, \mathrm{T}^{<\beta}) \ ,$$

then

$$\langle \widehat{\mathrm{U}}_{a,b}, \widehat{\mathrm{U}}_{a',b'}\rangle \in \mathrm{V}^\alpha$$

and we define

$$\mathrm{T}^\alpha(\widehat{\mathrm{U}}_{a,b}) := \mathrm{V}^{<\beta}$$

for the minimal β, which has the mentioned closure properties.

Note that with this construction we obtain that

$$\begin{array}{c} \mathrm{V}^\alpha \in \mathrm{L}'_{\alpha+1} \ , \\ \forall x \in \mathrm{Flat}(\mathrm{V}^\alpha).\mathrm{T}^\alpha(x) \in \mathrm{L}'_\alpha \ . \end{array}$$

Let HI be the least recursively hyper-inaccessible. Then one interprets

$$\begin{array}{rcl} [\![\,\mathrm{V}\,]\!] & := & \mathrm{V}^{<\mathrm{HI}} \ , \\ [\![\,\mathrm{T_V}\,]\!] & := & \mathrm{T}^{<\mathrm{HI}} \ , \\ [\![\,\mathrm{U}_{a,b}\,]\!]_\rho & := & \mathrm{T}^{<\mathrm{HI}}([\![\,\widehat{\mathrm{U}}_{a,b}\,]\!]_\rho) \ . \end{array}$$

Note that this model construction won't admit elimination rules for $\mathrm{U}_{a,b}$, since it is not interpreted as the least set closed under a, b and the basic set constructions – the $\mathrm{V}^{\beta'}$ chosen above might contain elements of the form $\widehat{\mathrm{U}}_{a',b'}$, which have been added, since the model of V needs to be closed under $\widehat{\mathrm{U}}$, but which are not elements of the least subuniverse closed under a, b and the standard set constructions. One could modify the above model in order to admit elimination rules: Then one would try to form at each stage α the least subset of $\mathrm{V}^{<\alpha}$ closed under the universe constructions and containing a and $b(x)$ for $x \in \mathrm{Flat}(\mathrm{T}^{<\alpha}(a))$. If this definition succeeds and the least such set is A, one would add $\widehat{\mathrm{U}}_{a,b}$ to V^α and define $\mathrm{T}^\alpha(\widehat{\mathrm{U}}_{a,b}) := A$.

We have to verify that $\mathrm{V}^{<\mathrm{HI}}$ has the necessary closure properties of the super universes. Since HI is a recursively inaccessible, $\mathrm{V}^{<\mathrm{HI}}$ will be closed under universe constructions as before. $\mathrm{V}^{<\mathrm{HI}}$ is as well closed under $(a, b)\widehat{\mathrm{U}}_{a,b}$: Assume

$$\langle\langle a, b\rangle, \langle a', b'\rangle\rangle \in [\![\,\mathrm{Fam}\,]\!](\mathrm{V}^{<\mathrm{HI}}, \mathrm{T}^{<\mathrm{HI}}) \ .$$

By the admissibility of HI it follows by Lemma 3.2

$$\langle\langle a, b\rangle, \langle a', b'\rangle\rangle \in [\![\,\mathrm{Fam}\,]\!](\mathrm{V}^{<\alpha}, \mathrm{T}^{<\alpha})$$

for some $\alpha < \mathrm{HI}$ Since HI is a recursively hyper inaccessible, there exists a recursively inaccessible ordinal $\alpha < \kappa < \mathrm{HI}$. V^κ will be closed under the universe constructions and

$$\langle\langle a, b\rangle, \langle a', b'\rangle\rangle \in [\![\, \mathrm{Fam} \,]\!](\mathrm{V}^\kappa, \mathrm{T}^\kappa) \ ,$$

therefore

$$\langle \widehat{\mathrm{U}}_{a,b}, \widehat{\mathrm{U}}_{a',b'}\rangle \in \mathrm{V}^{\kappa+2} \subseteq \mathrm{V}^{<\mathrm{HI}} \ .$$

That $[\![\, \mathrm{U}_{a,b} \,]\!]_\rho$ has the closure properties needed in order to interpret $\mathrm{U}_{a,b}$ follows by the construction.

The remaining construction is as for any standard model of type theory, i.e. we interpret the constructions of the small logical framework and the basic set constructions on top of U. As for $\mathrm{ML_W U}$, we need $< \omega$ many admissibles in order to interpret W-types built on top of $\mathrm{V}^{<\mathrm{HI}}$, therefore finite approximations of the type theory can be interpret into Kripke Platek set theory plus one recursively hyper-inaccessible and finitely many admissibles above it. So we have given the essence of a proof of the following theorem:

Theorem 4.2.

(a) *We can model* $\mathrm{ML_W} + (\mathrm{Superuniv})$ *in* $(\mathrm{KP} + (\mathrm{hyper} - \mathrm{inacc}))^+$.

(b) $|\mathrm{ML_W} + (\mathrm{Superuniv})| \leq |(\mathrm{KP} + (\mathrm{hyper} - \mathrm{inacc}))^+|$

(c) *The previous statements hold as well if we replace intensional by extensional equality.*

In order to obtain the precise proof theoretic strength, one has to carry out a well-ordering proof in order to obtain

$$|\mathrm{ML_W} + (\mathrm{Superuniv})| \geq |(\mathrm{KP} + (\mathrm{hyper} - \mathrm{inacc}))^+| \ .$$

We note that if one adds as in Erik's original type theory the full universe operator $\mathrm{SU} : \mathrm{Fam}(\mathrm{Set}) \to \mathrm{Fam}(\mathrm{Set})$, we get a little bit more strength. Formulated without the logical framework, in this type theory one has depending on $A : \mathrm{Set}$ and $x : A \Rightarrow B : \mathrm{Set}$ the set $\mathrm{U}(A, (x)B) : \mathrm{Set}$ and decoding function

$$u : \mathrm{U}(A, (x)B) \Rightarrow \mathrm{T}(A, (x)B, u) : \mathrm{Set} \ .$$

$(\mathrm{U}(A, (x)B), \mathrm{T}(A, (x)B))$ is a universe containing codes for A and $B[x := a]$, and in the above definition one replaces

$$\begin{array}{rll} \mathrm{U}_{a,b} & \text{by} & \mathrm{U}(\mathrm{T_V}(a), (x)\mathrm{T_V}(b(x))) \ , \\ \mathrm{T}_{a,b}(c) & \text{by} & \mathrm{T}(\mathrm{T_V}(a), (x)\mathrm{T_V}(b(x))) \ . \end{array}$$

In this theory one can apply the super universe operator as well to V itself, and to elements formed from it. However, only finite nestings of the super universe operator are possible. The modelling of one application of the super universe to arguments (A, B) requires one recursively inaccessible above $\max\{\mathrm{o}(A), \mathrm{o}(B)\}$. The modelling of V requires one recursively hyper-inaccessible, so in total one needs for Palmgren's original type theory one recursively hyper-inaccessible ordinal and finitely many recursively inaccessible ordinals above it.

5 The Mahlo Universe and Extensions of It

5.1 Definition of the Mahlo Universe

Once one has defined the super universe, one can move to a hyper-universe by forming a universe U^2 with decoding function T^2, s.t. for any family of sets a, b in U^2 there exists a super universe $\mathrm{U}^1_{a,b}$ and an element $\widehat{\mathrm{U}}^1_{a,b}$ in V, s.t. $\mathrm{T_V}(\widehat{\mathrm{U}}^1_{a,b}) = \mathrm{U}^1_{a,b}$, where $\mathrm{U}^1_{a,b}$ contains codes which decode as a, b as before. So we have a decoding function $\widehat{\mathrm{T}}^1_{a,b} : \mathrm{U}^1_{a,b} \to \mathrm{U}^2$, and define $\mathrm{T}^1_{a,b} := \mathrm{T}^2 \circ \widehat{\mathrm{T}}^1_{a,b}$. $\mathrm{U}^1_{a,b}$ is closed under the standard universe constructions. That $\mathrm{U}^1_{a,b}$ is a super universe means, that for any family of sets c, d in $\mathrm{U}^1_{a,b}$ there exists a subuniverse $\mathrm{U}^2_{a,b,c,d}$ of $\mathrm{U}^1_{a,b}$ which contains codes decoding as c, d, is closed under the standard universe constructions, and which is represented as a code in $\mathrm{U}^1_{a,b}$.

One could define now hyper$^{\alpha}$-super universes, but we won't spell this out in this article. The construction is similar to the step from a Mahlo-universe to a hyper$^{\alpha}$-Mahlo universe, which will be discussed in the followup of this article [32].

We note that E. Palmgren has [20] introduced higher order universes. In those universes, the lowest level is a universe, the next level is a universe of universe operators, the next level is a universes of operators on universe operators, etc. E. Palmgren added rules expressing the closure of higher order universes under the application of a universe operator from the next level. It is conjectured that E. Palmgren's universe reaches the strength of KPM, which is slightly below the strength of the Mahlo universe. We will not investigate this construction any further (see the remark in the introduction that Palmgren's universe contains a Mahlo universe in disguise, even so it is not easy to see this).

We will investigate in which sense the hyper$^{\alpha}$-super universes are special instances of universes closed under operators on families of sets, and in which sense the Mahlo universe to be introduced generalises this.

Definition 5.1. *(Assumes the logical framework).*

(a) An operator on families of sets is a function $f : \mathrm{Fam(Set)} \to \mathrm{Fam(Set)}$.

(b) Let $\langle \mathrm{U}, \mathrm{T} \rangle : \mathrm{Fam(Set)}$. *An operator on families of elements of* U *is a function* $f : \mathrm{Fam(U, T)} \to \mathrm{Fam(U, T)}$.

(c) Let $\langle A, B \rangle, \langle A', B' \rangle : \mathrm{Fam(Set)}$. *Assume elimination of* $+$ *into* Set. *Then*

$$\langle A, B \rangle \cup_{\mathrm{Fam(Set)}} \langle A', B' \rangle := \langle C, D \rangle$$

where

$$C := (A + \mathrm{N}_1) + (A' + \mathrm{N}_1)$$

and D is such that (remember that A^1_0 *is the canonical element of* N_1*)*

$$\begin{array}{ll} D(\mathrm{inl}(\mathrm{inl}(a))) = B(a) \ , & D(\mathrm{inl}(\mathrm{inr}(\mathrm{A}^1_0))) = A \ , \\ D(\mathrm{inr}(\mathrm{inl}(a'))) = B'(a') \ , & D(\mathrm{inr}(\mathrm{inr}(\mathrm{A}^1_0))) = A' \ . \end{array}$$

The super universe operator SU takes A : Fam(Set) and forms a universe closed under the operator $(X)A$: Fam(Set) $\to$ Fam(Set). The super universe is a universe which is closed under the super universe operator SU. One can define now a hyper-universe operator HU, which takes A : Fam(Set) and forms a super universe closed $(X)A$. One obtains essentially the same set by taking a universe closed under $(X)\mathrm{SU}(X) \cup_{\mathrm{Fam(Set)}} A$ (assuming this is definable). Then one can form a hyper^2-universe operator, a universe closed under it etc.

The step towards the Mahlo universe is now to form a universe, which is closed under arbitrary operators on families of elements of itself. This means that we have a universe $(\mathrm{V}, \mathrm{T}_{\mathrm{V}})$, s.t. for any operator $f : \mathrm{Fam}(\mathrm{V}, \mathrm{T}_{\mathrm{V}}) \to \mathrm{Fam}(\mathrm{V}, \mathrm{T}_{\mathrm{V}})$ there exists a subuniverse $(\mathrm{U}_f, \mathrm{T}_f)$ of V, closed under f and represented as an element $\widehat{\mathrm{U}}_f$ in V.

Spelled out without using the logical framework, we obtain the following rules:

- We have the rules of the small logical framework.
- We have the rules of the basic set constructions.
- We have rules expressing that $(\mathrm{V}, \mathrm{T}_{\mathrm{V}})$ is a universe closed under the basic set constructions. The constructors for the codes for these set constructions will be denoted by $\widehat{\mathrm{N}}_{\mathrm{V}}$, $\widehat{\Sigma}_{\mathrm{V}}$, etc.
- We have that for every operator f on families of sets in $(\mathrm{V}, \mathrm{T}_{\mathrm{V}})$ there is a a subuniverse U_f closed under it. We split f into two components and uncurry it, and obtain the following rules:

$$\frac{f : (a : \mathrm{V}, b : \mathrm{T}_{\mathrm{V}}(a) \to \mathrm{V}) \to \mathrm{V} \qquad g : (a : \mathrm{V}, b : \mathrm{T}_{\mathrm{V}}(a) \to \mathrm{V}, \mathrm{T}_{\mathrm{V}}(f(a,b)) \to \mathrm{V}}{\mathrm{U}_{f,g} : \mathrm{Set}}$$

- We have rules expressing that $\mathrm{U}_{f,g}$ is a recursive subuniverse closed under the standard set constructions. So we have the decoding function

$$\widehat{\mathrm{T}}_{\mathrm{U},f,g} : \mathrm{U}_{f,g} \to \mathrm{V}$$

  and define

$$\mathrm{T}_{\mathrm{U},f,g}(a) := \mathrm{T}_{\mathrm{V}}(\widehat{\mathrm{T}}_{\mathrm{U},f,g}(a)) : \mathrm{Set} \ .$$

  The codes for the basic set constructions will be denoted by $\widehat{\mathrm{N}}_{\mathrm{U},f,g}$, $\widehat{\Sigma}_{\mathrm{U},f,g}$, etc.

- We have rules expressing that $(\mathrm{U}_{f,g}, \mathrm{T}_{f,g})$ is closed under f and g:

$$\begin{array}{rcl}
\widehat{\mathrm{f}}_{f,g} & : & (a : \mathrm{U}_{f,g}, b : \mathrm{T}_{\mathrm{U},f,g}(a) \to \mathrm{U}_{f,g}) \to \mathrm{U}_{f,g} \\
\widehat{\mathrm{T}}_{\mathrm{U},f,g}(\widehat{\mathrm{f}}_{f,g}(a,b)) & = & f(\widehat{\mathrm{T}}_{\mathrm{V}}(a), \widehat{\mathrm{T}}_{\mathrm{V}} \circ b). \\
\widehat{\mathrm{g}}_{f,g} & : & (a : \mathrm{U}_{f,g}, b : \mathrm{T}_{\mathrm{U},f,g}(a) \to \mathrm{U}_{f,g}, \mathrm{T}_{\mathrm{V}}(f(\widehat{\mathrm{T}}_{\mathrm{V}}(a), \widehat{\mathrm{T}}_{\mathrm{V}} \circ b))) \to \mathrm{U}_{f,g} \\
\widehat{\mathrm{T}}_{\mathrm{U},f,g}(\widehat{\mathrm{g}}_{f,g}(a,b,c)) & = & g(\widehat{\mathrm{T}}_{\mathrm{V}}(a), \widehat{\mathrm{T}}_{\mathrm{V}} \circ b, c).
\end{array}$$

- And finally, we have the rule that V contains a code for $\mathrm{U}_{f,g}$. As P. Martin-Löf observed this is the rule which makes the Mahlo universe so strong – having just subuniverses closed under f, g doesn't add any strength to $\mathrm{ML_W U}$.

$$\frac{f : (a : \mathrm{V}, b : \mathrm{T_V}(a) \to \mathrm{V}) \to \mathrm{V} \qquad g : (a : \mathrm{V}, b : \mathrm{T_V}(a) \to \mathrm{V}, \mathrm{T_V}(f(a,b)) \to \mathrm{V}}{\begin{array}{c}\widehat{\mathrm{U}}_{f,g} : \mathrm{V} \\ \widehat{\mathrm{T}}_{\mathrm{V}}(\widehat{\mathrm{U}}_{f,g}) = \mathrm{U}_{f,g}\end{array}}$$

We call the resulting type theory $\mathrm{ML_W}$ + (Mahlo).

5.2 A Model of the Mahlo Universe

$\mathrm{ML_W}$ + (Mahlo) has strength $(\mathrm{KPM})^+$, where $(\mathrm{KPM})^+$ is $\mathrm{KP}\omega$ plus the existence of one recursively Mahlo set plus finitely many admissibles above it, which is equivalent to KPl plus the existence of one recursively Mahlo set.

Here a *recursively Mahlo set* is an admissible set $\mathrm{a_M}$ s.t.

$$\begin{array}{ll} \textit{(Mahlo)} & \forall \vec{z} \in \mathrm{a_M}.((\forall x \in \mathrm{a_M}.\exists y \in \mathrm{a_M}.\varphi(x,y,\vec{z})) \\ & \quad \to \exists z \in \mathrm{a_M}.\mathrm{Ad}(z) \wedge \vec{z} \in \mathrm{a_M}. \wedge (\forall x \in z.\exists y \in z.\varphi(x,y,\vec{z}))) \\ & \text{where } \varphi(x,y,\vec{z}) \text{ is } \Delta_0 \end{array}$$

A *recursively Mahlo ordinal* is an ordinal M s.t. $\mathrm{L_M}$ is a recursively Mahlo ordinal.

The basic construction of the Mahlo universe is as for the super universe: We define by recursion over α approximations V^α of $[\![\,\mathrm{V}\,]\!]$ by closing it systematically under the basic set constructions, and by adding suitable codes $\widehat{\mathrm{U}}_{f,g}$ to it.

The question is when to add codes $\widehat{\mathrm{U}}_{f,g}$ to the Mahlo universe. In the introduction rules, f, g need to be total functions from families of elements in V into families of elements in V. But we only know this type of total functions, once the definition of $[\![\,\mathrm{V}\,]\!]$ is complete. The trick to get around this problem is to add more elements $\widehat{\mathrm{U}}_{f,g}$ to the universe than are actually justified by the introduction rules. We only demand that f, g are total on the subuniverse $\mathrm{U}_{f,g}$ itself. So we take f, g to be arbitrary terms. Then we try to form a subuniverse of the current approximation of V^α of $[\![\,\mathrm{V}\,]\!]$, which is closed under f and g. If we succeed, we add $\widehat{\mathrm{U}}_{f,g}$ to $[\![\,\mathrm{V}\,]\!]$ and define $[\![\,\mathrm{T_V}\,]\!](\widehat{\mathrm{U}}_{f,g})$ as the least such subuniverse.

If we assume now total functions f, g from families of elements of $\mathrm{V}^{<\mathrm{M}}$ into families of elements of $\mathrm{V}^{<\mathrm{M}}$, then $\mathrm{V}^{<\mathrm{M}}$ is closed under f, g, where M is a recursively Mahlo ordinal. So there is at least one subuniverse of $\mathrm{V}^{<\mathrm{M}}$ closed under f, g, namely $\mathrm{V}^{<\mathrm{M}}$ itself. Then we will use the fact that M is a Mahlo-ordinal, and find that such a universe occurred already at some stage $\alpha < \mathrm{M}$, and that therefore $\widehat{\mathrm{U}}_{f,g}$ is in $\mathrm{V}^{<\mathrm{M}}$.

As for the super universe, we simplify this construction slightly and don't search for the least subuniverse closed under f and g, but only for the least β s.t. V^β is closed under f and g.

More precisely we proceed as follows: As in the model for the super universe, we forget in the terms of the model the typing information needed in $\widehat{\mathrm{N}}$, $\widehat{\Sigma}$, etc. (but not in $\widehat{\mathrm{f}}$, $\widehat{\mathrm{g}}$), so for instance

$$[\![\widehat{\mathrm{N}}_{\mathrm{V}}]\!]_\rho \quad := \quad [\![\widehat{\mathrm{N}}_{\mathrm{U},f,g}]\!]_\rho := \widehat{\mathrm{N}} \enspace .$$

We obtain essentially the same reduction rules for $\widehat{\mathrm{T}}_{\mathrm{U},f,g}$ as the reduction rules for $\widehat{\mathrm{T}}_{\mathrm{U},a,b}$ in case of the super universe. Furthermore we have the reductions

$$\begin{array}{rcl} \widehat{\mathrm{f}}_{f,g}(a,b) & \longrightarrow & f(a,b) \enspace , \\ \widehat{\mathrm{g}}_{f,g}(a,b,c) & \longrightarrow & g(a,b,c) \enspace . \end{array}$$

As for the super universe we form $(\mathrm{V}^\alpha, \mathrm{T}^\alpha) \in [\![\mathrm{Fam}]\!](\mathrm{set})$ by closing it under the standard set constructions. We will then interpret $[\![\mathrm{V}]\!] = \mathrm{V}^{<\mathrm{M}}$, $[\![\mathrm{T}_{\mathrm{V}}]\!] = \mathrm{T}^{<\mathrm{M}}$.

Instead of closing it under $\widehat{\mathrm{U}}_{a,b}$ for $\langle a, b\rangle \in [\![\mathrm{Fam}]\!](\mathrm{V}^{<\alpha}, \mathrm{T}^{<\alpha})$ we close it under $\widehat{\mathrm{U}}_{f,g}$, where f, g form the two components of a function

$$[\![\mathrm{Fam}]\!](\mathrm{V}^{<\alpha}, \mathrm{T}^{<\alpha}) \; [\![\rightarrow]\!] \; [\![\mathrm{Fam}]\!](\mathrm{V}^{<\alpha}, \mathrm{T}^{<\alpha}) \enspace ,$$

and we call $\langle f, g\rangle$ an operator on families of sets in $(\mathrm{V}^{<\alpha}, \mathrm{T}^{<\alpha})$. More precisely, we introduce the following definition, where we first define a notion for the two components of an operator on families of sets in $\langle A, B\rangle$, and then define what it means to be an operator on families of such sets as the conjunction of the two:

Definition 5.2. *Let* $\langle A, B\rangle \in [\![\mathrm{Fam}]\!](\mathrm{set})$.

(a) $$\begin{array}{rl} [\![\mathrm{FamOper}]\!]_0(A, B) := & (x \in A) \\ & [\![\rightarrow]\!]((y \in B(x)) \; [\![\rightarrow]\!] \; A) \\ & [\![\rightarrow]\!] A \enspace . \end{array}$$

(b) Assume $f \in \mathrm{Flat}([\![\mathrm{FamOper}]\!]_0(A, B))$. *Then*

$$\begin{array}{rl} [\![\mathrm{FamOper}]\!]_1(A, B, f) := & (x \in A) \\ & [\![\rightarrow]\!]((y \in B(x)) \; [\![\rightarrow]\!] \; A) \\ & [\![\rightarrow]\!] B(f(x, y)) \\ & [\![\rightarrow]\!] A \enspace . \end{array}$$

(c) The set of operators on families of sets in $\langle A, B\rangle$ *is given as*

$$\begin{array}{l} [\![\mathrm{FamOper}]\!](A, B) := \\ \quad \{\langle\langle f, g\rangle, \langle f', g'\rangle\rangle \in ([\![\mathrm{Term}]\!]^2)^2 \mid \langle f, f'\rangle \in [\![\mathrm{FamOper}]\!]_0(A, B) \wedge \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad \langle g, g'\rangle \in [\![\mathrm{FamOper}]\!]_1(A, B, f)\} \enspace . \end{array}$$

With this notation, the addition of $\widehat{\mathrm{U}}_{f,g}$ to V^α is defined as follows:

Assume $f, f', g, g' \in [\![\mathrm{Term}]\!]$. Assume $\alpha < \mathrm{M}$ and β s.t. $\beta + 1 < \alpha$ and such that $\mathrm{V}^{<\beta}$ is closed under all universe constructions. Assume that

$$\langle\langle f, g\rangle, \langle f', g'\rangle\rangle \in [\![\mathrm{FamOper}]\!](\mathrm{V}^{<\beta}, \mathrm{T}^{<\beta})$$

Let β be minimal that this property holds. Then

$$\langle \widehat{\mathrm{U}}_{f,g}, \widehat{\mathrm{U}}_{f',g'} \rangle \in \mathrm{V}^{\alpha} \quad ,$$

and define

$$\mathrm{T}^{\alpha}(\widehat{\mathrm{U}}_{f,g}) := \mathrm{V}^{<\beta} \quad .$$

Note that this construction preserves the property that

$$\mathrm{V}^{\alpha} \in \mathrm{L}'_{\alpha+1} \wedge \forall a \in \mathrm{Flat}(\mathrm{V}^{\alpha}).\mathrm{T}^{\alpha}(a) \in \mathrm{L}'_{\alpha} \quad .$$

Now define

$$\begin{array}{rcl} [\![\,\mathrm{U}\,]\!]_{\rho} & := & \mathrm{V}^{<\mathrm{M}} \quad , \\ [\![\,\mathrm{T}_{\mathrm{V}}(c)\,]\!]_{\rho} & := & \mathrm{T}^{<\mathrm{M}}([\![\,c\,]\!]_{\rho}) \quad , \\ [\![\,\mathrm{U}_{f,g}\,]\!]_{\rho} & := & \mathrm{T}^{<\mathrm{M}}(\widehat{\mathrm{U}}_{[\![\,f\,]\!]_{\rho},[\![\,g\,]\!]_{\rho}}) \quad . \end{array}$$

where M is the least recursively Mahlo ordinal.

We note that $\mathrm{T}^{\alpha}(\widehat{\mathrm{U}}_{f,g})$ is not defined as a minimal set closed under universe constructions, and under f and g: We could have done so by modifying the above definition and trying at stage α to define the least subset of V^{α} closed under f and g and the standard universe constructions. If this succeeds, this subset is independent of α (provided α is big enough so that this subset lies inside V^{α}) and we could have interpreted $\mathrm{T}^{\alpha}(\widehat{\mathrm{U}}_{f,g})$ as this set. This would interpret elimination rules for $\mathrm{U}_{f,g}$ as well. But since elimination rules for $\mathrm{U}_{f,g}$ don't add any proof theoretic strength, and that model construction is more complicated, we use the original approach.

We have to verify that $\mathrm{V}^{<\mathrm{M}}$ has the necessary closure properties of the Mahlo universe. Every recursively Mahlo ordinal is recursively inaccessible, therefore $\mathrm{V}^{<\mathrm{M}}$ will be closed under universe constructions as before. $\mathrm{V}^{<\mathrm{M}}$ is as well closed under $(f,g)\widehat{\mathrm{U}}_{f,g}$: Assume $f, f', g, g' \in [\![\,\mathrm{Term}\,]\!]$. Assume

$$\begin{array}{rcl} \langle f, f' \rangle & \in & [\![\,(x : \mathrm{V}, y : \mathrm{T}_{\mathrm{V}}(x) \to \mathrm{V}) \to \mathrm{V}\,]\!] \quad , \\ \langle g, g' \rangle & \in & [\![\,(x : \mathrm{V}, y : \mathrm{T}_{\mathrm{V}}(x) \to \mathrm{V}, z(\mathrm{T}_{\mathrm{V}}(x), \mathrm{T}_{\mathrm{V}} \circ y)) \to \mathrm{V}\,]\!]_{[z=f]} \quad . \end{array}$$

This means

$$\langle \langle f, g \rangle, \langle f', g' \rangle \rangle \in [\![\,\mathrm{FamOper}\,]\!](\mathrm{V}^{<\mathrm{M}}, \mathrm{T}^{<\mathrm{M}})$$

We need to find a $\kappa < \mathrm{M}$ s.t. $\mathrm{V}^{<\kappa}$ is closed under universe constructions and that the previous conditions hold with M replaced by κ. For being closed under universe constructions it suffices that κ is inaccessible. We show that for every $\beta < \mathrm{M}$ there exists a $\rho < \mathrm{M}$ s.t. ρ is admissible, $\beta < \rho$ and s.t.

$$\begin{array}{l} \langle \langle a, b \rangle, \langle a', b' \rangle \rangle \in [\![\,\mathrm{Fam}\,]\!](\mathrm{V}^{<\beta}, \mathrm{T}^{<\beta}) \\ \quad \to \langle \langle f(a,b), g(a,b) \rangle, \langle f'(a',b'), g'(a',b') \rangle \rangle \in [\![\,\mathrm{Fam}\,]\!](\mathrm{V}^{<\rho}, \mathrm{T}^{<\rho}) \end{array}$$

This formula will then be expressed by a Π_2-formula and by the Mahloness of M we then find a κ s.t. for $\beta < \kappa$ there exists a $\rho < \kappa$ which fulfils the previous conditions. Then κ will be inaccessible and $\mathrm{V}^{<\kappa}$ will be closed under $\langle f, f' \rangle$ and $\langle g, g' \rangle$.

So assume $\beta < \mathrm{M}$ and

$$\langle\langle a, b\rangle, \langle a', b'\rangle\rangle \in [\![\,\mathrm{Fam}\,]\!](\mathrm{V}^{<\beta}, \mathrm{T}^{<\beta}) \ .$$

Then there exists a γ s.t. $\langle f(a, b), f'(a', b')\rangle \in \mathrm{V}^{<\gamma}$. Furthermore,

$$\mathrm{T}^{<\mathrm{M}}(f(a, b)) = \mathrm{T}^{<\gamma}(f(a, b)) \in \mathrm{L}'_{\gamma} \ ,$$

and for $\langle c, c'\rangle \in \mathrm{T}^{<\mathrm{M}}(f(a, b))$ there exists a $\delta < \mathrm{M}$ s.t.

$$\langle g(a, b, c), g'(a', b', c')\rangle \in \mathrm{T}^{<\delta}(f(a, b)) \ .$$

By the admissibility of M it follows that there exists a $\rho' < \mathrm{M}$ s.t. for $\langle c, c'\rangle \in \mathrm{T}^{<\mathrm{M}}(f(a, b))$ we have

$$\langle g(a, b, c), g'(a', b', c')\rangle \in \mathrm{T}^{<\rho'}(f(a, b)) \ .$$

We can obtain that $\gamma \leq \rho'$ and therefore

$$\langle\langle f(a, b), f'(a', b')\rangle, \langle g(a, b), g'(a', b')\rangle\rangle \in [\![\,\mathrm{Fam}\,]\!](\mathrm{V}^{<\rho'}, \mathrm{T}^{<\rho'}) \ .$$

Using again the admissibility of M and the fact that $[\![\,\mathrm{Fam}\,]\!](\mathrm{V}^{<\rho'}, \mathrm{T}^{<\rho'}) \in \mathrm{L}'_{\rho'+2}$ we obtain depending on β a uniform ρ s.t. if

$$\langle\langle a, b\rangle, \langle a', b'\rangle\rangle \in [\![\,\mathrm{Fam}\,]\!](\mathrm{V}^{<\beta}, \mathrm{T}^{<\beta})$$

then

$$\langle\langle f(a, b), f'(a', b')\rangle, \langle g(a, b), g'(a', b')\rangle\rangle \in [\![\,\mathrm{Fam}\,]\!](\mathrm{V}^{<\rho}, \mathrm{T}^{<\rho}) \ .$$

Furthermore, using the fact that every recursively Mahlo ordinal is recursively inaccessible, we can achieve $\beta < \rho$ admissible. Therefore, for every $\beta < \mathrm{M}$ there exists a $\beta < \rho < \mathrm{M}$ s.t. ρ is admissible and s.t. if

$$\langle\langle a, b\rangle, \langle a', b'\rangle\rangle \in [\![\,\mathrm{Fam}\,]\!](\mathrm{V}^{<\beta}, \mathrm{T}^{<\beta}) \ ,$$

then

$$\langle\langle f(a, b), g(a, b)\rangle, \langle f'(a', b'), g'(a', b')\rangle\rangle \in [\![\,\mathrm{Fam}\,]\!](\mathrm{V}^{<\rho}, \mathrm{T}^{<\rho}) \ .$$

The property we have shown can be expressed as a Π_2-formula, i.e. a formula of the form

$$\forall\beta < \mathrm{M}.\exists\rho < \mathrm{M}.\varphi(\beta, \rho) \ ,$$

where φ is Δ_0.

By the Mahlo property we obtain the existence of an admissible $\kappa < \mathrm{M}$ s.t.

$$\forall\beta < \kappa.\exists\rho < \kappa.\varphi(\beta, \rho) \ .$$

Since ρ is always an admissible $> \beta$, it follows that κ is recursively inaccessible, therefore $(\mathrm{V}^{<\kappa}, \mathrm{T}^{<\kappa})$ is closed under the basic set constructions. Furthermore if

$$\langle\langle a, b\rangle, \langle a', b'\rangle\rangle \in [\![\,\mathrm{Fam}\,]\!](\mathrm{V}^{<\kappa}, \mathrm{T}^{<\kappa}) \ ,$$

then

$$\langle\langle a,b\rangle,\langle a',b'\rangle\rangle \in [\![\, \mathrm{Fam} \,]\!](\mathrm{V}^{<\beta},\mathrm{T}^{<\beta})$$

for some $\beta < \kappa$. Therefore there exists a $\rho < \kappa$ s.t.

$$\langle\langle f(a,b),g(a,b)\rangle,\langle f'(a',b'),g'(a',b')\rangle\rangle \in [\![\, \mathrm{Fam} \,]\!](\mathrm{V}^{<\rho},\mathrm{T}^{<\rho}) \subseteq [\![\, \mathrm{Fam} \,]\!](\mathrm{V}^{<\kappa},\mathrm{T}^{<\kappa}) \ .$$

This means that $\mathrm{V}^{<\kappa}$ is a candidate for the interpretation of $\langle \widehat{\mathrm{U}}_{f,g}, \widehat{\mathrm{U}}_{f',g'}\rangle$, and therefore

$$\langle \widehat{\mathrm{U}}_{f,g}, \widehat{\mathrm{U}}_{f',g'}\rangle \in \mathrm{V}^{\kappa+2} \subseteq \mathrm{V}^{<\mathrm{M}} \ .$$

That $[\![\, \mathrm{U}_{f,g} \,]\!]_\rho$ has the closure properties needed in order to interpret $\mathrm{U}_{f,g}$ follows by the construction.

The remaining steps are as for the model of simple $\mathrm{ML_W U}$, i.e. we need $< \omega$ many admissibles in order to interpret the basic set constructions on top of $\mathrm{V}, \mathrm{T_V}$. Therefore the type theory can be interpreted in Kripke Platek set theory plus one recursively Mahlo ordinal and finitely many admissibles above it. So we have given the essence of a proof of the following theorem:

Theorem 5.3.

(a) We can model $\mathrm{ML_W} + (\mathrm{Mahlo})$ *in* $(\mathrm{KPM})^+$.

(b) $|\mathrm{ML_W} + (\mathrm{Mahlo})| \leq |(\mathrm{KPM})^+|$

(c) The previous statements hold as well if we replace intensional by extensional equality.

By [30] we obtain the other direction and obtain therefore

Theorem 5.4. *(a)* $|\mathrm{ML_W} + (\mathrm{Mahlo})| = |(\mathrm{KPM})^+|$

(b) The previous statements hold as well if we replace intensional by extensional equality.

References

[1] P. Aczel. The strength of Martin-Löf's intuitionistic type theory with one universe. In S. Miettinen and J. Väänänen, editors, *Proceedings of the symposium on Mathematical Logic (Oulu 1974)*. Report No. 2, Dept. of Philosophy, Univ of Helsinki, 1977.

[2] T. Arai. Ordinal diagrams for Π_3-reflection. *J. Symbolic Logic*, 65(3):1375 – 1394, 2000.

[3] T. Arai. Ordinal diagrams for recursively Mahlo universes. *Arch. Math. Logic*, 39(5):353 – 391, 2000.

[4] T. Arai. Proof theory for theories of ordinals. I. Recursively Mahlo ordinals. *Ann. Pure Appl. Logic*, 122(1 – 3):1 – 85, 2003.

[5] T. Arai. Proof theory for theories of ordinals. II. Π_3-reflection. *Ann. Pure Appl. Logic*, 129(1 – 3):39 – 92, 2004.

[6] J. Barwise. *Admissible Sets and Structures. An Approach to Definability Theory.* Omega-series. Springer, 1975.

[7] M. Benke, P. Dybjer, and P. Jansson. Universes for generic programs and proofs in dependent type theory. *Nordic J. of Computing*, 10(4):265–289, 2003.

[8] P. Dybjer. A general formulation of simultaneous inductive-recursive definitions in type theory. *Journal of Symbolic Logic*, 65(2):525–549, June 2000.

[9] P. Dybjer and A. Setzer. Finite axiomatizations of inductive and inductive-recursive definitions. In *Workshop on Generic Programming, Marstrand, Sweden, 18 June 1998.* http://www.cs.ruu.nl/people/johanj/programme_wgp98.html, 1998.

[10] P. Dybjer and A. Setzer. A finite axiomatization of inductive-recursive definitions. In J.-Y. Girard, editor, *Typed Lambda Calculi and Applications*, volume 1581 of *Lecture Notes in Computer Science*, pages 129–146. Springer, April 1999.

[11] P. Dybjer and A. Setzer. Indexed induction-recursion. In R. Kahle, P. Schroeder-Heister, and R. Stärk, editors, *Proof Theory in Computer Science*, volume 2183 of *Lecture Notes in Computer Science*, pages 93 – 113. Springer, 2001.

[12] P. Dybjer and A. Setzer. Induction-recursion and initial algebras. *Annals of Pure and Applied Logic*, 124:1–47, 2003.

[13] P. Dybjer and A. Setzer. Indexed induction-recursion. 64 pp. Submitted., 2005.

[14] E. Griffor and M. Rathjen. The strength of some Martin-Löf type theories. *Arch. math. Logic*, 33:347 – 385, 1994.

[15] M. Hofmann. Syntax and semantics of dependent types. In A. M. Pitts and P. Dybjer, editors, *Semantics and logics of computation*, pages 79 – 130, Cambridge, 1997. Cambridge University Press.

[16] G. Jäger. *Theories for Admissible Sets: A Unifying Approach to Proof Theory.* Bibliopolis, Naples, 1986.

[17] P. Martin-Löf. *Intuitionistic type theory.* Bibliopolis, Naples, 1984.

[18] B. Nordström, K. Petersson, and J. Smith. *Programming in Martin-Löf's type theory. An Introduction.* Oxford University-Press, Oxford, 1990. Out of print. Online version available via http://www.cs.chalmers.se/Cs/Research/Logic/book/.

[19] B. Nordström, K. Petersson, and J. M. Smith. Martin-löf's type theory. In S. Abramsky, D. M. Gabbay, and T. S. E. Maibaum, editors, *Handbook of logic in computer science, Vol. 5*, pages 1 – 37. Oxford Univ. Press, 2000.

[20] E. Palmgren. On universes in type theory. In G. Sambin and J. Smith, editors, *Twenty five years of constructive type theory*, pages 191 – 204, Oxford, 1998. Oxford University Press.

[21] A. Ranta. *Type-theoretical Grammar.* Clarendon Press, Oxford, 1995.

[22] M. Rathjen. Proof-theoretical analysis of KPM. *Arch. math. Logic*, 30:377 – 403, 1991.

[23] M. Rathjen. Proof theory of reflection. *Ann. Pure Appl. Logic*, 68:181 – 224, 1994.

[24] M. Rathjen. Recent advances in ordinal analysis: Π^1_2-CA and related systems. *Bulletin of Symbolic Logic*, 1:468 – 485, 1995.

[25] A. Setzer. *Proof theoretical strength of Martin-Löf Type Theory with W-type and one universe.* PhD thesis, Universität München, 1993. available from http://www.cs.swan.ac.uk/∼csetzer.

[26] A. Setzer. A model for a type theory with Mahlo Universe. 10pp. Preprint, available via http://www.cs.swan.ac.uk/∼/articles/uppermahlo.pdf, 1996.

[27] A. Setzer. An upper bound for the proof theoretical strength of Martin-Löf Type Theory with W-type and one Universe. 33pp. Draft, available via http://www.cs.swan.ac.uk/∼/articles/1papdiss.dvi or .ps or .pdf, 1996.

[28] A. Setzer. An introduction to well-ordering proofs in Martin-Löf's type theory. In G. Sambin and J. Smith, editors, *Twenty-five years of constructive type theory*, pages 245 – 263, Oxford, 1998. Clarendon Press.

[29] A. Setzer. Well-ordering proofs for Martin-Löf type theory. *Annals of Pure and Applied Logic*, 92:113 – 159, 1998.

[30] A. Setzer. Extending Martin-Löf type theory by one Mahlo-universe. *Arch. Math. Log.*, 39:155 – 181, 2000.

[31] A. Setzer. Proof theory of Martin-Löf type theory – an overview. *Mathematiques et Sciences Humaines*, 42 année, n°165:59 – 99, 2004.

[32] A. Setzer. Universes in type theory part II – Π_3-reflection. In preparation, 2005.

[33] S. G. Simpson. *Subsystems of second-order arithmetic.* Springer, 1999.

[34] A. Troelstra and D. v. Dalen. *Constructivism in Mathematics. An Introduction, Vol. II.* North-Holland, Amsterdam, 1988.