UNIVERSES OVER FREGE STRUCTURES

R. KAHLE

REPORT No. 20, 2000/2001
ISSN 1103-467X
ISRN IML-R- -20-00/01- -SE

INSTITUT MITTAG-LEFFLER
THE ROYAL SWEDISH ACADEMY OF SCIENCES

Universes over Frege structures

Reinhard Kahle

WSI, Universität Tübingen,
Sand 13, D-72076 Tübingen, Germany
e-mail: kahle@informatik.uni-tuebingen.de

March 31, 2001

Abstract

In this paper we study a concept of universe for a truth predicate over applicative theories. A proof-theoretic analysis is given by use of transfinitely iterated fixed point theories $\widehat{\mathsf{ID}}_\alpha$. The lower bound is obtained by a syntactical interpretation of these theories. Thus, universes over Frege structures represent a syntactically expressive framework of metapredicative theories in the context of applicative theories.

1 Introduction

Based on the prior work of Scott [Sco75] *Frege structures* were introduced by Aczel as a semantical concept to define a notion of set on the bases of a *partial truth predicate* [Acz80]. Frege structures can be axiomatized as a truth theory over *applicative theories* [Acz77, Bee85]. These theories build the first-order part of Feferman's systems of explicit mathematics [Fef75, Fef79].

Applicative theories are studied in detail at several places from different points of view, cf. [Bee85, JS95b, Str96, Kah97a, JKS99]. In general, they comprise partial combinatory logic and axioms for additional constants, at least pairing and projection, and natural numbers.

Frege structures can by axiomatized by adding a truth predicate to applicative theories. Compared with other truth theories the applicative basis has the great advantage that we need not do any form of Gödelization. Instead, we just add new terms to the language to represent formulae. Truth theories for the applicative framework are studied at length in Cantini's monograph [Can96]. One specific topic is the analysis of iterated truth predicates, cf. also [Can95]. Our approach is closely related to his formalism. However, while Cantini uses external terms for the levels of truth, our iteration is based on a defined relation within the applicative language.

Using Aczel's original idea and defining an element relation in terms of the truth predicate, the closure conditions of the truth levels can be read as set theoretical closure conditions similar to those for universes in Martin-Löf's type theory or in explicit mathematics [Mar84, Mar94]. The so-called *uniform limit axiom* of our theory allows us to build transfinite hierarchies of universes. For this reason, we get a very smooth syntactical treatment of the metapredicative, transfinitely iterated fixed point theories $\widehat{\mathsf{ID}}_\alpha$. In fact, our theories were one of the first metapredicative theories and they were the reason for the proof-theoretic analysis of $\widehat{\mathsf{ID}}_\alpha$ in [JKSS99]. In the meanwhile, metapredicativity became a flourishing branch of proof theory, cf. [JKSS99, Set99, Str99, Rat00, Rüe00, Str00a, Rat0x, Jäg0x, JS0xa, Rüe0xa, Rüe0xb]. In particular, the concept of superuniverse in Martin-Löf's type theory, studied by Rathjen [Rat00, Rat0x], and metapredicative versions of explicit mathematics with universes are closely related to our approach. The latter ones are studied by Strahm [Str99] and, in fact, his work and the presented one here were discussed in parallel and have profit from each other.

The structure of the paper is as follows. In the next section, we introduce the applicative theory TON. Then, we present the truth theory FON with its main results, especially the syntactical embedding of $\widehat{\mathsf{ID}}_1$. In section 4, we define the theory FSU of universes over Frege structures and prove the existence of hierarchies of universes which allow us to define fixed points for appropriate operator forms. The last section is devoted to the proof theory of FSU.

2 The theory TON

The theory TON (total theory of operations and numbers) introduced and studied by Jäger and Strahm [JS95b], is the total version of the theory BON (basic theory of operations and numbers) introduced of Feferman and Jäger [FJ93]. It is formulated in $\mathcal{L}_t$, the first order language of operation and numbers. $\mathcal{L}_t$ comprises individual variables $x, y, z, u, v, w, f, g, h, \ldots$, individual constants k, s (combinators), p, p_0, p_1 (pairing and projection), 0, $\mathsf{s_N}$, $\mathsf{p_N}$ (zero, successor and predecessor), $\mathsf{d_N}$ (definition by cases), a binary function symbol $\cdot$ for term application, and the relation symbols $=$ and N. Terms $(r, s, t, \ldots)$ are built up from individual variables and individual constants by term application. Formulae $(\varphi, \psi, \ldots)$ are defined by $\neg$, $\wedge$ and $\forall$ as usual, starting from the atomic formulae $t = s$ and $\mathsf{N}(t)$.

In the following, we write $s\,t$ for $(s \cdot t)$ with the convention of association to the left. The connectives $\vee$, $\to$ and $\exists$ are defined as usual. Quantifiers which are restricted to elements of N are written in the form of $\forall x : \mathsf{N}.\varphi$.

The logic of TON is classical first-order predicate logic with equality. The non-logical axioms of TON include:

I. **Combinatory algebra.**

(1) $\mathsf{k}\,x\,y = x$,

(2) $\mathsf{s}\,x\,y\,z = x\,z\,(y\,z)$,

II. **Pairing and projection.**

(3) $\mathsf{p}_0\,(\mathsf{p}\,x\,y) = x \wedge \mathsf{p}_1\,(\mathsf{p}\,x\,y) = y$,

III. **Natural numbers.**

(4) $\mathsf{N}(0) \wedge \forall x.\mathsf{N}(x) \to \mathsf{N}(\mathsf{s_N}\,x)$,

(5) $\forall x.\mathsf{N}(x) \to \mathsf{s_N}\,x \neq 0 \wedge \mathsf{p_N}\,(\mathsf{s_N}\,x) = x$,

(6) $\forall x.\mathsf{N}(x) \wedge x \neq 0 \to \mathsf{N}(\mathsf{p_N}\,x) \wedge \mathsf{s_N}\,(\mathsf{p_N}\,x) = x$.

IV. **Definition by cases on N.**

(7) $\mathsf{N}(v) \wedge \mathsf{N}(w) \wedge v = w \to \mathsf{d_N}\,x\,y\,v\,w = x$,

(8) $\mathsf{N}(v) \wedge \mathsf{N}(w) \wedge v \neq w \to \mathsf{d_N}\,x\,y\,v\,w = y$.

The definition of applicative theories in [FJ93] and [JS95b] also comprises primitive recursion. It is needed only for very restricted forms of induction which we do not discuss here.

In the standard way, we can introduce the usual λ notation in TON. Moreover, the type-free approach allows self-application such that we can define a recursion operator rec.

Proposition 1

1. For every variable x and every term t of $\mathcal{L}_t$, there exists a term $\lambda x.t$ of $\mathcal{L}_t$ whose free variables are those of t, excluding x, such that TON proves $(\lambda x.t)\,x = t$.

2. There exists a term rec of $\mathcal{L}_t$ such that TON proves $\forall x.\mathsf{rec}\,x = x\,(\mathsf{rec}\,x)$.

The theory TON is proof-theoretically equivalent to Peano arithmetic PA if we add the schema of formulae induction on N:

Formulae induction on N ($\mathcal{L}_t$-$\mathrm{I_N}$)

$$\varphi(0) \land (\forall x : \mathsf{N}.\varphi(x) \to \varphi(\mathsf{s_N}\, x)) \to \forall x : \mathsf{N}.\varphi(x).$$

By interpreting natural numbers as elements of N, we get a straightforward interpretation of PA in $\mathsf{TON} + (\mathcal{L}_t\text{-}\mathrm{I_N})$, cf. [JS95b].

Models of TON are discussed at length by Beeson [Bee85], Cantini [Can96] and Strahm [JS95b, Str96]. Later on, our model construction for universes will be based on the *closed total term model* $\mathcal{CTT}$ which can be roughly described as follows:

The universe of $\mathcal{CTT}$ consists of the set of all closed terms of the language $\mathcal{L}_t$, i.e. we interpret the constants by themselves, and application by juxtaposition. Based on a straightforward reduction relation ϱ for the constants of $\mathcal{L}_t$, we interpret equality of terms by the fact that they have a common reduct with respect to *arbitrary* reductions on the basis of ϱ. For this reason, we need to prove the *Church-Rosser property* to verify transitivity of equality. $\mathsf{N}(t)$ holds if t reduces to a numeral.

For more details about $\mathcal{CTT}$ we refer to [JS95b, Can96], in particular for the proof of the Church-Rosser property.

This model can be formalized in PA. Thus, we get the proof-theoretic equivalence of $\mathsf{TON} + (\mathcal{L}_t\text{-}\mathrm{I_N})$ and PA.

In the sequel we need some ordinal-theoretic facts as given in §2 of [JKSS99]. In particular, we have to extend the standard notation system for ordinals up to the Feferman-Schütte ordinal Γ_0 which is based on the well-known binary Veblen function $\boldsymbol{\varphi}$, cf. [Poh89] and [Sch77]. Therefore, we introduce the ternary Veblen function as follows:

1. $\boldsymbol{\varphi}\, 0\, \beta\, \gamma := \boldsymbol{\varphi}\, \beta\, \gamma$,
2. for $\alpha > 0$, $\boldsymbol{\varphi}\, \alpha\, 0\, \gamma$ denotes the γth ordinal which is strongly critical with respect to all functions $\lambda\xi, \eta.\boldsymbol{\varphi}\, \alpha'\, \xi\, \eta$ for $\alpha' < \alpha$,
3. for $\alpha > 0$ and $\beta > 0$, $\boldsymbol{\varphi}\, \alpha\, \beta\, \gamma$ denotes the γth common fixed point of the functions $\lambda\xi.\boldsymbol{\varphi}\, \alpha\, \beta'\, \xi$ for $\beta' < \beta$.

Given a standard notation system based on the ternary φ function, we write $\prec$ for the corresponding primitive recursive wellordering. Of course, there exist primitive recursive functions representing the usual ordinal operations like plus, times, exponentiation and $\boldsymbol{\varphi}$. Within the formal theories we use lower Greek letters $\alpha, \beta, \gamma, \ldots$ as variables ranging over the field of $\prec$. For limit notations, we use Λ (since λ is already used for the λ-notation in the calculus). We do not distinguish explicitly between the ordinals with their ordinal operations and their codes and primitive recursive analogues. As abbreviations we use $\forall\alpha \prec \beta.\varphi$ for $\forall\alpha.\alpha \prec \beta \to \varphi$ and $\exists\alpha \prec \beta.\varphi$ for $\exists\alpha.\alpha \prec \beta \land \varphi$. For limit notations, we use Λ (since λ is already used for the λ-notation in the calculus). Finally, we define as usual:

$$\begin{aligned} \mathrm{Prog}(\varphi) \quad &:\Leftrightarrow \quad \forall\gamma.(\forall\beta \prec \gamma.\varphi(\beta)) \to \varphi(\gamma), \\ \mathrm{TI}(\alpha, \varphi) \quad &:\Leftrightarrow \quad \mathrm{Prog}(\varphi) \to \forall\beta \prec \alpha.\varphi(\beta). \end{aligned}$$

3 Frege structures

Aczel's concept of Frege structures [Acz80] were axiomatized by Beeson as a truth theory over applicative theories [Bee85]. Here, we define Frege structures in a very general way, like the truth theories of Cantini in [Can93, Can96]. The relation to the various variants which can be found in the literature is discussed at the end of this section.

The theory FON (Frege structures over TON) is formulated in the language $\mathcal{L}_F$, which expands $\mathcal{L}_t$ by a unary relation symbol T (truth) and individual constants $\doteq$, $\dot{\mathsf{N}}$, $\dot{\neg}$, $\dot{\wedge}$ and $\dot{\forall}$.

For the sake of readability in the following, we use freely an infix notation for terms containing the dotted constants. For instance, $x \doteq y$ has to be written formally as $\doteq x\, y$.

The axioms of FON are those of TON extended to the expanded language, as well as the following ones:

I. **Closure under prime formulae of TON**

(1) $x = y \leftrightarrow \mathsf{T}(x \doteq y)$,

(2) $\neg x = y \leftrightarrow \mathsf{T}(\dot{\neg}\,(x \doteq y))$,

(3) $\mathsf{N}(x) \leftrightarrow \mathsf{T}(\dot{\mathsf{N}}\, x)$,

(4) $\neg \mathsf{N}(x) \leftrightarrow \mathsf{T}(\dot{\neg}\,(\dot{\mathsf{N}}\, x))$.

II. **Closure under composed formulae**

(5) $\mathsf{T}(x) \leftrightarrow \mathsf{T}(\dot{\neg}\,(\dot{\neg}\, x))$,

(6) $\mathsf{T}(x) \wedge \mathsf{T}(y) \leftrightarrow \mathsf{T}(x \mathbin{\dot{\wedge}} y)$,

(7) $\mathsf{T}(\dot{\neg}\, x) \vee \mathsf{T}(\dot{\neg}\, y) \leftrightarrow \mathsf{T}(\dot{\neg}\,(x \mathbin{\dot{\wedge}} y))$,

(8) $(\forall x.\mathsf{T}(f\, x)) \leftrightarrow \mathsf{T}(\dot{\forall}\, f)$,

(9) $(\exists x.\mathsf{T}(\dot{\neg}\,(f\, x))) \leftrightarrow \mathsf{T}(\dot{\neg}\,(\dot{\forall}\, f))$.

III. **Consistency**

(10) $\neg(\mathsf{T}(x) \wedge \mathsf{T}(\dot{\neg}\, x))$

With respect to the well-known problems of truth definitions, we cannot add full self-reference in the sense that there is a term $\dot{\mathsf{T}}$ such that $\neg\mathsf{T}(x) \leftrightarrow \mathsf{T}(\dot{\neg}\,(\dot{\mathsf{T}}\, x))$. Nevertheless, we get a rather trivial form of self-reference:

Self-reference

$$\mathsf{T}(x) \leftrightarrow \mathsf{T}(\dot{\mathsf{T}}\, x),$$
$$\mathsf{T}(\dot{\neg}\, x) \leftrightarrow \mathsf{T}(\dot{\neg}\,(\dot{\mathsf{T}}\, x)),$$

by defining $\dot{\mathsf{T}}$ as the identity function $\lambda x.x$.

As abbreviations we introduce the following notions:

$$\begin{aligned} \mathsf{F}(t) &:\Leftrightarrow \mathsf{T}(\dot{\neg}\, t), \\ \mathsf{P}(t) &:\Leftrightarrow \mathsf{T}(t) \vee \mathsf{F}(t), \\ \mathsf{C}(t) &:\Leftrightarrow \forall x.\mathsf{P}(t\, x). \end{aligned}$$

$\mathsf{T}(t)$ can be read as "t is true", $\mathsf{F}(t)$ as "t is false", $\mathsf{P}(t)$ as "t is a proposition", and $\mathsf{C}(t)$ as "t is a class". Classes are also called "propositional functions". By use of the recursion theorem, we can diagonalize $\dot{\neg}$ and get that there are objects which are not propositions.

Lemma 2 $\mathsf{FON} \vdash \neg\forall x.\mathsf{P}(x)$.

Proof: We set $l := \mathsf{rec}\,(\lambda x.\dot\neg\, x)$, i.e. $l = \dot\neg\, l$. So l can be seen as a formalization of the liar sentence. Now, $\mathsf{T}(l)$ is equivalent with $\mathsf{T}(\dot\neg\, l)$. Thus $\mathsf{P}(l)$ would contradict the axiom of consistency. $\square$

Frege structures allow us to represent $\mathcal{L}_F$ formulae by terms without any form of Gödelization: In a straightforward manner, we define for every $\mathcal{L}_F$-formula φ a term $\dot\varphi$ by replacing $=$, N, T, $\neg$, and $\wedge$ by the corresponding individual constants $\dot=$, $\dot{\mathsf{N}}$, $\dot{\mathsf{T}}$, $\dot\neg$, and $\dot\wedge$, respectively. In the case of the universal quantifier, we set $\dot{\overbrace{\forall x.\varphi}} := \dot\forall\,(\lambda x.\dot\varphi)$. Because we work in classical logic, we can introduce the other logical constants as abbreviations on the term level, too: $(t \mathbin{\dot\vee} s) := \dot\neg\,(\dot\neg\, t \mathbin{\dot\wedge} \dot\neg\, s)$, $(t \mathbin{\dot\rightarrow} s) := \dot\neg\, t \mathbin{\dot\vee} s$ and $(\dot\exists x.t) := \dot\neg\,(\dot\forall x.\dot\neg\, t)$. Using the "dot"-notation $\dot\varphi$ we would like to indicate the difference of this representation compared with a Gödelization $\ulcorner\varphi\urcorner$. From a technical point of view, the representation is particular convenient with respect to variables and substitution. Since terms and variables stay untouched in the term representing a formula, we have $\dot\varphi[t/x] \equiv \dot{\overbrace{\varphi[t/x]}}$.

With the positive part of self-reference, our axiomatization provides a truth definition for T-positive formulae "only". We define the notion of T-positive formulae simultaneously with T-negative formulae:

Definition 3

1. $t = s$, $\mathsf{N}(t)$, $\neg t = s$ and $\neg\mathsf{N}(t)$ are T-positive as well as T-negative.
2. $\mathsf{T}(t)$ is T-positive; $\neg\mathsf{T}(t)$ is T-negative.
3. If φ is T-positive (T-negative), then $\neg\varphi$ is T-negative (T-positive).
4. If φ and ψ are T-positive (T-negative), then so is $\varphi \wedge \psi$.
5. If φ is T-positive (T-negative), then so is $\forall x.\varphi$.

By induction on the structure of formulae we get the following crucial result for Frege structures [Can96, theorem 8.8.]:

Proposition 4 If φ is a T-positive formulae of $\mathcal{L}_F$, then we have:

$$\mathsf{FON} \vdash \mathsf{T}(\dot\varphi) \leftrightarrow \varphi.$$

Based on this proposition, we can introduce a notion of *set* and define an element-relation:

Definition 5 Given two $\mathcal{L}_F$ terms t and s and a $\mathcal{L}_F$ formula φ, we define:

$$\begin{aligned} \{x|\varphi\} &:= \lambda x.\dot\varphi, \\ t \in s &:\Leftrightarrow \mathsf{T}(s\,t). \end{aligned}$$

Please note that the expression $\{x|\varphi\}$ is defined for arbitrary formulae φ. In this sense, we can say that Frege structures allow full or unrestricted comprehension. On the other hand, only in the case of T-positive formulae, we get as a corollary from proposition 4 the intended behavior:

Corollary 6 If φ is a T-positive formulae of $\mathcal{L}_F$, then we have:

$$\mathsf{FON} \vdash x \in \{x|\varphi\} \leftrightarrow \varphi.$$

As induction principles, we will consider the following three forms of induction over N.

1. Class induction on N (C-I_N)

$$\mathsf{C}(f) \wedge \mathsf{T}(f\,0) \wedge (\forall x : \mathsf{N}.\mathsf{T}(f\,x) \to \mathsf{T}(f\,(\mathsf{s_N}\,x))) \to \forall x : \mathsf{N}.\mathsf{T}(f\,x).$$

2. Truth induction on N (T-I_N)

$$\mathsf{T}(f\,0) \wedge (\forall x : \mathsf{N}.\mathsf{T}(f\,x) \to \mathsf{T}(f\,(\mathsf{s_N}\,x))) \to \forall x : \mathsf{N}.\mathsf{T}(f\,x).$$

3. Formulae induction on N ($\mathcal{L}_F$-I_N)

$$\varphi(0) \wedge (\forall x : \mathsf{N}.\varphi(x) \to \varphi(\mathsf{s_N}\,x)) \to \forall x : \mathsf{N}.\varphi(x).$$

With respect to our definition of sets, the first two induction principles can also be written as follows:

1′. Class induction on N (C-I_N)

$$\mathsf{C}(f) \wedge 0 \in f \wedge (\forall x : \mathsf{N}.x \in f \to \mathsf{s_N}\,x \in f) \to \forall x : \mathsf{N}.x \in f.$$

2′. Truth induction on N (T-I_N)

$$0 \in f \wedge (\forall x : \mathsf{N}.x \in f \to \mathsf{s_N}\,x \in f) \to \forall x : \mathsf{N}.x \in f.$$

Essentially, the theories $\mathsf{FON}, \mathsf{FON} + (\text{C-}\mathrm{I}_\mathsf{N})$, $\mathsf{FON} + (\text{T-}\mathrm{I}_\mathsf{N})$, and $\mathsf{FON} + (\mathcal{L}_F\text{-}\mathrm{I}_\mathsf{N})$ are equivalent with the theories MF^-, MF_c, MF_p, and MF, respectively, of Cantini in [Can96].

In [Sco75] Scott has defined a fixed point of sets by use of a fixed point theorem on the ground structure. In the same manner, Cantini proved that the truth theories can be used to define *fixed points* of positive operator forms, [Can96, §10A]. Let us define the notion of positive operator form in FON as a T-positive formula $\varphi(R, x)$ which contains the fresh (unary) relation variable R only positively.

Proposition 7 Let $\varphi(R, x)$ be a positive operator form. Then there exists a term t_φ of $\mathcal{L}_F$ such that

$$\mathsf{FON} \vdash \forall x.\mathsf{T}(t_\varphi\,x) \leftrightarrow \varphi(\mathsf{T}(t_\varphi\,\cdot), x).$$

Proof: We define t_φ by use of the recursion theorem.

$$t_\varphi \quad := \quad \mathsf{rec}\,(\lambda y, x.\overset{\cdot}{\overbrace{\varphi(\mathsf{T}(y\,\cdot), x)}})$$

So we have $t_\varphi\,x = \overset{\cdot}{\overbrace{\varphi(\mathsf{T}(t_\varphi\,\cdot), x)}}$ and get

$$\begin{aligned} \mathsf{T}(t_\varphi\,x) \quad &\leftrightarrow \quad \mathsf{T}(\overset{\cdot}{\overbrace{\varphi(\mathsf{T}(t_\varphi\,\cdot), x)}}) \\ &\leftrightarrow \quad \varphi(\mathsf{T}(t_\varphi\,\cdot), x). \end{aligned}$$

Of course, the T-positiveness of φ is crucial to apply proposition 4. □

This proposition allows us to give a *syntactical* interpretation of fixed point theories $\widehat{\mathsf{ID}}_1$ in $\mathsf{FON} + (\mathcal{L}_F\text{-}\mathrm{I}_\mathsf{N})$, and of its restriction $\mathsf{ID}_1^\#$ in $\mathsf{FON} + (\text{T-}\mathrm{I}_\mathsf{N})$.

For the fixed point theories $\widehat{\mathsf{ID}}_1$ and $\mathsf{ID}_1^\#$, we extend Peano arithmetic by (not necessarily least) fixed points $\mathcal{P}^\varphi$ of P-positive arithmetical operator forms $\varphi(P, x)$, i.e. an arithmetical formula in which $P(t)$ occurs only positively, cf. [BFPS81].

Definition 8 The language $\mathcal{L}_{\mathsf{ID}}$ of $\widehat{\mathsf{ID}}_1$ and $\mathsf{ID}_1^{\#}$ is the language of PA extended by new fixed point constants $\mathcal{P}^{\varphi}$ for each P-positive arithmetical operator form $\varphi(P, x)$.

For both theories, we extend the axioms of PA by the following fixed point axiom for each fixed point constants $\mathcal{P}^{\varphi}$:

$$\forall x.\varphi(\mathcal{P}^{\varphi}, x) \leftrightarrow \mathcal{P}^{\varphi}(x).$$

While we have induction on the natural numbers for all formulae of the extended language in $\widehat{\mathsf{ID}}_1$, in $\mathsf{ID}_1^{\#}$ the induction schema is restricted to formulae which have only positive occurrences of the new fixed point constants.

By use of proposition 4, we can carry over the embedding of PA in $\mathsf{TON} + (\mathcal{L}_t\text{-}\mathrm{I}_{\mathsf{N}})$ to an embedding of it in $\mathsf{FON} + (\mathsf{C}\text{-}\mathrm{I}_{\mathsf{N}})$. This interpretation can be extended to an embedding of $\widehat{\mathsf{ID}}_1$ and $\mathsf{ID}_1^{\#}$ in Frege structures by use of proposition 7.

Proposition 9 There exists a translation ^{N} from the language $\mathcal{L}_{\mathsf{ID}}$ in the language $\mathcal{L}_F$ such that

1. $\mathsf{ID}_1^{\#} \vdash \varphi \quad \Rightarrow \quad \mathsf{FON} + (\mathsf{T}\text{-}\mathrm{I}_{\mathsf{N}}) \vdash \varphi^N$,
2. $\widehat{\mathsf{ID}}_1 \vdash \varphi \quad \Rightarrow \quad \mathsf{FON} + (\mathcal{L}_F\text{-}\mathrm{I}_{\mathsf{N}}) \vdash \varphi^N$.

Proof: We extend the standard interpretation ^{N} of PA in $\mathsf{TON} + (\mathcal{L}_t\text{-}\mathrm{I}_{\mathsf{N}})$ by interpreting the new fixed point constants $\mathcal{P}^{\varphi}(x)$ by $\mathsf{T}(t_{\varphi^N}\, x)$ where t_{φ^N} is defined as in the proof of proposition 7. Thus, P-positive operator forms become positive operator forms in FON and the proposition above verifies the additional fixed point axioms.

Of course, $(\mathcal{L}_F\text{-}\mathrm{I}_{\mathsf{N}})$ verifies the formulae induction of $\widehat{\mathsf{ID}}_1$. Since the induction principle of $\mathsf{ID}_1^{\#}$ allows only positive occurrences of the fixed point constants which are interpreted by T-positive formulae, the translation of the induction schema can be proven by $(\mathsf{T}\text{-}\mathrm{I}_{\mathsf{N}})$. □

For the proof of the upper bounds, we refer to Cantini [Can96, §57]. From his work, we get the following theorem about the proof-theoretic equivalences (for the meaning of proof-theoretic equivalence denoted by $\equiv$, we refer to [Fef88] and [Fef00]).

Theorem 10

1. $\mathsf{FON} + (\mathsf{C}\text{-}\mathrm{I}_{\mathsf{N}}) \equiv \mathsf{PA}$,
2. $\mathsf{FON} + (\mathsf{T}\text{-}\mathrm{I}_{\mathsf{N}}) \equiv \mathsf{ID}_1^{\#}$,
3. $\mathsf{FON} + (\mathcal{L}_F\text{-}\mathrm{I}_{\mathsf{N}}) \equiv \widehat{\mathsf{ID}}_1$.

The standard model construction for our truth theory can be given by a straightforward operator form over a given model $\mathcal{M}$ of TON, cf. [FM87b, §4]. We assume that M is the universe of $\mathcal{M}$ containing distinct elements $\dot{=}$, $\dot{\neg}$, $\dot{\wedge}$, $\dot{\mathsf{N}}$ and $\dot{\forall}$. Moreover, we will identify the terms of $\mathcal{L}_F$ with their interpretation in $\mathcal{M}$. Let $\mathcal{N}$ be the interpretation of N. Now, we define an operator $\mathcal{T}(a)$ which holds if there exist elements b, c and f such that one of the following clauses holds.

1. $a = (b \mathbin{\dot{=}} c) \wedge b = c$,
2. $a = \dot{\neg}\,(b \mathbin{\dot{=}} c) \wedge b \neq c$,
3. $a = \dot{\mathsf{N}}\, b \wedge \mathcal{N}(b)$,
4. $a = \dot{\neg}\,(\dot{\mathsf{N}}\, b) \wedge \neg\mathcal{N}(b)$,
5. $a = \dot{\neg}\,(\dot{\neg}\, b) \wedge \mathcal{T}(b)$,
6. $a = (b \mathbin{\dot{\wedge}} c) \wedge \mathcal{T}(b) \wedge \mathcal{T}(c)$,

7. $a = \dot\neg\,(b \dot\wedge c) \wedge (\mathcal{T}(\dot\neg\, b) \vee \mathcal{T}(\dot\neg\, c))$,

8. $a = \dot\forall f \wedge \forall x.\mathcal{T}(f\, x)$,

9. $a = \dot\neg\,(\dot\forall f) \wedge \exists x.\mathcal{T}(\dot\neg\,(f\, x))$.

To handle consistency we have to single out the "consistent" elements of $\mathcal{T}$. Therefore, we can set $\mathcal{T}^+(a) :\Leftrightarrow \mathcal{T}(a) \wedge \neg\mathcal{T}(\dot\neg\, a)$ and using $\mathcal{T}^+$ for the interpretation of the truth predicate T. Now, the verification of the axioms of FON is straightforward.

The operator $\mathcal{T}$ can be easily formalized in $\widehat{\mathsf{ID}}_1$ such that the clauses 1.–9. are satisfied. However, because of the negative occurrence of $\mathcal{T}$ in the definition of $\mathcal{T}^+$ we can not verify consistency directly in $\widehat{\mathsf{ID}}_1$. Of course, it will follow in the theory ID_1 of non-iterated inductive definitions which extends $\widehat{\mathsf{ID}}_1$ with a minimality condition (induction principle) for the fixed points, cf. [Poh89]. With respect to proof-theoretic strength, ID_1 is much too strong. Cantini proved by methods of ordinal analysis that $\mathsf{FON} + (\mathcal{L}_F\text{-}\mathrm{I}_{\mathsf{N}})$ indeed does not exceed the strength of $\widehat{\mathsf{ID}}_1$. Let us give a sketch how consistency can be handled along a proof-theoretic treatment of fixed point theories with ordinals. Such theories, introduced by Jäger [Jäg93], were widely used in the proof-theoretic analysis of explicit mathematics, cf. [FJ93, JS95b, JS95a, FJ96, GS96, JS96, MS98, FS00, Str00b]. For the analysis of $\mathsf{FON} + (\mathcal{L}_F\text{-}\mathrm{I}_{\mathsf{N}})$ we would proceed with the following steps.

1. Introducing a theory $\mathcal{T}^+\text{-}\widehat{\mathsf{ID}}_1$ consisting of an fixed point axiom for the (translation of the) operator $\mathcal{T}$ plus an additional axiom stating the consistency of the fixed point.

2. Defining a semi-formal system $\omega\text{-}\mathsf{PA}^w_\Omega$, i.e. the theory PA^w_Ω of [Jäg93] but involving a ω-rule for the natural numbers.

   The main feature of this theory is that it provides us with ordinals which allow to build fixed points in stages. Together with Δ^Ω_0-induction on the ordinals this can be used to verify the consistency of the fixed point in the next step.

3. Interpreting $\mathcal{T}^+\text{-}\widehat{\mathsf{ID}}_1$ in $\omega\text{-}\mathsf{PA}^w_\Omega$.

4. Proceed with the usual proof-theoretic methods in the analysis of $\omega\text{-}\mathsf{PA}^w_\Omega$: Partial cut elimination up to Σ^Ω/Π^Ω formulae for this theory; introducing a suitable infinitary system T_∞ which allows full cut-elimination and an asymmetric interpretation of $\omega\text{-}\mathsf{PA}^w_\Omega$.

Because the consistency was already treated in the third step by use of Δ^Ω_0-induction on the ordinals in $\omega\text{-}\mathsf{PA}^w_\Omega$, we will not exceed the proof-theoretic bounds as they are computed for an analysis of $\widehat{\mathsf{ID}}_1$ along these lines. The procedure described here is paradigmatic for the treatment of consistency as it can be done for the theories with universes later on.

A detailed work out of such a procedure can be found in [Str00b] where Strahm analyzes theories called $\mathsf{PA}^w_\Omega + (\mathsf{Subst})$ and $\mathsf{PA}^+_\Omega + (\mathsf{Subst})$. By replacing the Σ^Ω-induction by Δ^Ω_0-induction in the semi-formal system T which corresponds to PA^+_Ω one gets exactly the theory we called here $\omega\text{-}\mathsf{PA}^w_\Omega$. For the next step one can actually use exactly the semi-formal system T_∞ defined there, but replacing the bounds $(\beta, \varphi\,\alpha\,(\beta+\beta))$ in the assymetric interpretation by $(\beta, \beta+2^\alpha)$, cf. [Str00b, theorem 13].

We finish this section by a short discussion of our definition of Frege structures compared with other approaches.

First, in contrast to Beeson, we have defined the truth theory over the *total* version of applicative theories, because of a subtle problem with strictness in the partial setting. This problem is discussed in detail in [Kah99].

Secondly, Aczel's original definition of Frege structures (and its formalization by Beeson) presupposed an (additionally axiomatized) notion of *proposition* [Acz77, Acz80, Bee85]. Then the truth axioms are applicable only for arguments which are propositions. We follow the more general axiomatization as given by Flagg and Myhill [FM87a, FM87b]. In this context the notion of proposition is defined in terms of the truth predicate. In fact, this generalization is necessary to define fixed points in FON.

Hayashi and Kobayashi have given in [HK95] a version of Frege structures which is exactly tuned to be equivalent with a version of Feferman's explicit mathematics. This formalization uses propositions, too.

In general, based on the defined notion of sets, we can deal with (total versions of) theories of explicit mathematics, like the theory EET of Feferman and Jäger [FJ96], cf. [Can93, Can96]. But the converse does not hold. It is not known how one could define fixed points of arithmetical operator forms directly in theories of explicit mathematics.

Finally, the embedding of $\widehat{\mathsf{ID}}_1$ in $\mathsf{FON} + (\mathcal{L}_F\text{-}\mathrm{I}_{\mathsf{N}})$ makes essential use of the recursion, or fixed point theorem, on the term level. But there is no natural way, to define *least* fixed points on the term level. (For a definition of a least fixed point operator in applicative theories which makes essential use of *partiality* and additional *computability axioms*, cf. [KS0x].) Thus, we cannot expect to define a similar truth theory which allows an embedding of ID_1. However, Cantini has defined another truth theory over applicative theory based on *supervaluation* and proved its proof-theoretic equivalence with ID_1, cf. [Can96] (and [Can90] for a corresponding theory over PA). By use of his methods, this result can be slightly improved giving a syntactical interpretation of ID_1 in this theory, cf. [Kah97a, Kah0x].

4 Universes over Frege structures

The concept of *universes* was first introduced by Martin-Löf in his type theory, [Mar84]. Universes are types closed under the natural type existence operations. The proof-theoretic investigation of these theories led to the definition of (iterated) fixed point theories $\widehat{\mathsf{ID}}_1$ and $\widehat{\mathsf{ID}}_n$ by Aczel and Feferman, [Acz77, Fef82]. In the context of explicit mathematics, "predicative" universes were studied by Feferman [Fef82] and Marzetta [Mar93, Mar94]. Recently, Jäger, Kahle and Studer studied impredicative theories of explicit mathematics with universes, [JKS0x, JS0xb].

The definition of universes in Frege structures merges in some sense the ideas of universes in explicit mathematics with the concept of *iterated truth predicates.* In particular, Cantini's theory of TLR of *reflective truth with levels* is closely related to our approach, [Can96, chapter 8]. In this theory, truth is iterated by use of levels which are represented by a special sort of index terms. In contrast, our approach used a relation which is defined within the applicative language to iterate the concept of truth. For this reason, we stay completely in the applicative framework.

The existence of universes is guaranteed by a limit axiom, like for admissibles in Kripke-Platek set theory [Jäg84] or Marzetta's universes in explicit mathematics. In the latter theory, the limit axiom is *non-uniform* while its *uniform* version is studied by Kahle and Strahm [Kah97b, Str99]. Here, we will deal with the uniform version too, since it is essential for the definition of transfinite hierarchies of universes. For this reason we will get *metapredicative* theories, i.e. theories which exceed the proof-theoretic strength of predicativity, but which do not have essential impredicative features. In particular, its proof-theoretic analysis does not require tools of impredicative proof theory like collapsing, while yielding a proof-theoretic ordinal beyond Γ_0, the so-called limit of predicativity, cf. [Jäg0x].

The theory FSU

As mentioned above, there are two views of universes in Frege structures. On the one hand, they can be seen as *iterated truth "predicates"* (in fact they are terms, not predicates). This means that we can handle *negative* statements with respect to truth at a certain level as *positive* statements on a higher level. On the other hand, our universes are closed under natural set-theoretical operations.

This view is based on the "Janus face" of the truth predicates which provides us with an element relation, too.

Our universes will be classes in the sense defined above. All elements of such an universe are true, and universes are closed under the standard truth conditions. They will be ordered by use of a order relation $t \sqsubset s$. Essentially, this relation allows us to expresses negative statements about t as positive statements with respect to s. Formally, this relation is defined as follows:

Definition 11

$$t \sqsubset s \quad :\Leftrightarrow \quad \forall x.(\mathsf{T}(t\,x) \to \mathsf{T}(s\,(t\,x))) \wedge (\mathsf{T}(\dot\neg\,(t\,x)) \to \mathsf{T}(s\,(\dot\neg\,(t\,x)))).$$

If we extend the abbreviation $s \in t$ of $\mathsf{T}(t\,s)$ to negation and "punctuation", we can rephrase this definition in a more readable way. Let us define:

$$\begin{aligned} s \notin t \quad &:= \quad \mathsf{T}(\dot\neg\,(t\,s)) \\ \overset{\cdot}{\overbrace{s \in t}} \quad &:= \quad s \mathrel{\dot\in} t \quad := \quad t\,s \\ \overset{\cdot}{\overbrace{s \notin t}} \quad &:= \quad s \mathrel{\dot\notin} t \quad := \quad \dot\neg\,(t\,s). \end{aligned}$$

So we get:

$$t \sqsubset s \quad \Leftrightarrow \quad \forall x.(x \in t \to \overset{\cdot}{\overbrace{(x \in t)}} \in s) \wedge (x \notin t \to \overset{\cdot}{\overbrace{(x \notin t)}} \in s).$$

If $t \sqsubset s$ holds, s reflects, so to speak, the *truth-course-of-value* of the consistent elements of t, and we will say for short: s *reflects* t.

The existence of universes follows from the *limit axiom*: For each class, there exists a universe which *reflects* it. Moreover, the universes are introduced in a *uniform* way. This uniformity will be used later to build transfinite hierarchies of universes, whose length depends on the induction principle.

The theory FSU (Frege structures with universes) is formulated in the language $\mathcal{L}_U$, which expands $\mathcal{L}_F$ by the new relation symbol U and the new individual constant ℓ.

The axioms of FSU are those of TON extended to the expanded language and the following ones:

I. **Basic axioms**

(1) $\mathsf{U}(u) \to \mathsf{C}(u)$,

(2) $\mathsf{U}(u) \to \forall x. x \in u \to \mathsf{T}(x)$.

II. **Closure under prime formulae of TON**

(3) $\mathsf{U}(u) \to \forall x, y. x = y \leftrightarrow x \doteq y \in u$,

(4) $\mathsf{U}(u) \to \forall x, y. x \neq y \leftrightarrow \dot\neg\,(x \doteq y) \in u$,

(5) $\mathsf{U}(u) \to \forall x. \mathsf{N}(x) \leftrightarrow \dot{\mathsf{N}}\,x \in u$,

(6) $\mathsf{U}(u) \to \forall x. \neg\mathsf{N}(x) \leftrightarrow \dot\neg\,(\dot{\mathsf{N}}\,x) \in u$.

III. **Closure under composed formulae**

(7) $\mathsf{U}(u) \to \forall x. x \in u \leftrightarrow \dot\neg\,(\dot\neg\,x) \in u$,

(8) $\mathsf{U}(u) \to \forall x, y. x \in u \wedge y \in u \leftrightarrow x \mathbin{\dot\wedge} y \in u$,

(9) $\mathsf{U}(u) \to \forall x, y. \dot\neg\,x \in u \vee \dot\neg\,y \in u \leftrightarrow \dot\neg\,(x \mathbin{\dot\wedge} y) \in u$,

(10) $\mathsf{U}(u) \to ((\forall x. f\,x \in u) \leftrightarrow \dot\forall\,f \in u)$,

(11) $\mathsf{U}(u) \to ((\exists x.\dot\neg\,(f\,x) \in u) \leftrightarrow \dot\neg\,(\dot\forall\, f) \in u)$.

IV. **Order structure**

(12) $\mathsf{U}(u) \wedge \mathsf{U}(v) \wedge (t \mathrel{\dot\in} u) \in v \to t \in v$,

V. **Local consistency**

(13) $\mathsf{U}(u) \to \neg(x \in u \wedge \dot\neg\, x \in u)$,

VI. **Limit axiom**

(14) $\forall f.\mathsf{C}(f) \to \mathsf{U}(\ell\, f) \wedge f \sqsubset \ell\, f$.

As a first consequence we get that truth in general is equivalent with "truth in a universe" or "elementhood in a universe".

Lemma 12 $\mathsf{FSU} \vdash \mathsf{T}(x) \leftrightarrow \exists u.\mathsf{U}(u) \wedge x \in u$.

Proof: The direction from the right to the left is axiom I.(2). For the other one, it follows from $\mathsf{T}(x)$ that the constant function $\lambda y.x$ is a class. By applying the limit axiom we get $\exists u.\mathsf{U}(u) \wedge \lambda y.x \sqsubset u$. The definition of $\sqsubset$ and $\mathsf{T}(x)$ yield $x \in u$. □

The *global consistency* of T follows from local consistency.

Lemma 13 $\mathsf{FSU} \vdash \forall x.\neg(\mathsf{T}(x) \wedge \mathsf{T}(\dot\neg x))$.

Proof: Assume $\mathsf{T}(x) \wedge \mathsf{T}(\dot\neg x)$. Thus, $\lambda y.x$ is a class, and we can apply limit which yields: $\mathsf{U}(\ell\,(\lambda y.x)) \wedge (\lambda y.x) \sqsubset \ell\,(\lambda y.x)$. Now, the assumption together with the definition of $\sqsubset$ implies $x \in (\ell\,(\lambda y.x)) \wedge \dot\neg\, x \in (\ell\,(\lambda y.x))$ in contradiction to local consistency. □

By use of these two lemmas we can easily embed FON in FSU:

Proposition 14 Let φ be a $\mathcal{L}_F$ formula. Then we have:

$$\mathsf{FON} \vdash \varphi \quad \Rightarrow \quad \mathsf{FSU} \vdash \varphi.$$

Proof: The proof is by induction on the length of the derivation of φ in FON. The verification of the groups I. and II. of axioms of FON follows from lemma 12. Only the verification of closure for universal quantification needs an extra application of the limit axiom in the direction from the left to the right:

$$\begin{array}{rcl}
\forall x.\mathsf{T}(f\,x) & \to & \mathsf{C}(\lambda x.f\,x) \wedge \forall x.\mathsf{T}(f\,x) \\
& \to & \exists u.\mathsf{U}(u) \wedge (\lambda x.f\,x) \sqsubset u \wedge \forall x.\mathsf{T}(f\,x) \\
& \to & \exists u.\mathsf{U}(u) \wedge (\forall x.x \in f \to f\,x \in u) \wedge \forall x.x \in f \\
& \to & \exists u.\mathsf{U}(u) \wedge \forall x.f\,x \in u \\
& \to & \exists u.\mathsf{U}(u) \wedge \dot\forall\, f \in u \\
& \to & \mathsf{T}(\dot\forall\, f).
\end{array}$$

The axiom III.(10) of consistency follows from local consistency as shown in the previous lemma. □

The order relation $\sqsubset$ is an unbounded strict partial order on universes. In addition the relation is persistent (or transitive with respect to the defined element relation).

Lemma 15

1. $\mathsf{FSU} \vdash \mathsf{U}(u) \wedge \mathsf{U}(v) \wedge u \sqsubset v \wedge x \in u \to x \in v$,
2. $\mathsf{FSU} \vdash \mathsf{U}(u) \to \exists v.\mathsf{U}(v) \wedge u \sqsubset v$,
3. $\mathsf{FSU} \vdash \mathsf{U}(u) \to \neg u \sqsubset u$,
4. $\mathsf{FSU} \vdash \mathsf{U}(u) \wedge \mathsf{U}(v) \wedge \mathsf{U}(w) \wedge u \sqsubset v \wedge v \sqsubset w \to u \sqsubset w$.

Proof:

1. By definition of $\sqsubset$ we get from the antecedence $(x \mathrel{\dot\in} u) \in v$ which implies $x \in v$ by axiom IV.(12).
2. Since universes are classes, the assertion is a direct consequence of the limit axiom.
3. Let us define $r := \mathsf{rec}\,(\lambda x.x \mathrel{\dot\notin} u)$, i.e. $r = r \mathrel{\dot\notin} u$. Assuming $\mathsf{U}(u) \wedge u \sqsubset u$, we get the following contradiction:

$$\begin{array}{rcl} \mathsf{T}(u\,r) & \leftrightarrow & r \in u \\ & \leftrightarrow & (r \mathrel{\dot\notin} u) \in u \\ & \leftrightarrow & r \notin u \\ & \leftrightarrow & \neg\mathsf{T}(u\,r) \end{array}$$

4. Transitivity follows again from axiom IV.(12). Assume $\mathsf{U}(u) \wedge \mathsf{U}(v) \wedge \mathsf{U}(w) \wedge u \sqsubset v \wedge v \sqsubset w$. So we have

$$\begin{array}{rcl} x \in u & \to & (x \mathrel{\dot\in} u) \in v \\ & \to & ((x \mathrel{\dot\in} u) \mathrel{\dot\in} v) \in w \\ & \to & (x \mathrel{\dot\in} u) \in w \\ x \notin u & \to & (x \mathrel{\dot\notin} u) \in v \\ & \to & ((x \mathrel{\dot\notin} u) \mathrel{\dot\in} v) \in w \\ & \to & (x \mathrel{\dot\notin} u) \in w \end{array}$$

□

Let us finish this subsection with some remarks about FSU and related theories.

In view of the upper bounds we could strengthen the theories by adding directedness:

(U-Dir) $\quad \mathsf{U}(u) \wedge \mathsf{U}(v) \to \exists w.\mathsf{U}(w) \wedge w \sqsubset u \wedge v \sqsubset w$

or linearity of $\sqsubset$:

(U-Lin) $\quad \mathsf{U}(u) \wedge \mathsf{U}(v) \wedge \mathsf{U}(w) \to w \sqsubset u \vee u \equiv w \vee u \sqsubset w$.

Here, $u \equiv w$ stands for $\forall x.x \in u \leftrightarrow x \in w$. However, both is not needed for our purposes. For instance, in Cantini's theory TLR — which is closely related to FSU as we will see in a minute — directedness was used to prove global consistency, cf. [Can96, lemma 37.7.(i)]. In FSU the limit axiom together with the definition of $\sqsubset$ is enough to establish the corresponding result.

In the definition of the theory we have used a purely *axiomatization* of the notion of universes. Alternatively, it is possible to introduce this notion as an abbreviation. In this way, universes in explicit mathematics are treated e.g. in [JKS0x]. However, in this case one has to be careful with the possible ordering principles. It turns out that, for instance, linearity of the universes can only demanded for *normal universes*, i.e. universes generated (named) by the limit operation. The axiomatization, instead, allows easily to restrict the totality of universes. The model construction given below will even verify that every universe is normal in the following sense:

(U-Nor) $\quad \mathsf{U}(u) \to \exists f.\mathsf{C}(f) \wedge u = \ell\, f$.

Our theory FSU shares the main features of Cantini's theory TLR as described in [Can96, chapter 8]. This theory is defining truth levels by use of an external sort of level indexes. It can be roughly translated into FSU by translating the local truth predicates $\mathsf{T}_i(t)$ by $\mathsf{U}(u_i) \wedge \mathsf{T}(u_i\, t)$. The ordering on the levels, $i \prec j$ corresponds to our order relation $v \sqsubset u$. In addition, quantifiers for the level has to bound the corresponding universes in FSU. Since both, a formal definition of TLR, cf. [Can96, §36], as well as an exact proof of an interpretation in FSU would require some lengthy work, we just sketch the embedding. But, there are two limitation we have to notice. First, TLR contains directness of the levels. Therefore, a full interpretation would require the addition of (U-Dir) to FSU. However, it does not look like that directedness is needed in essential cases where it cannot be replaced by our limit axiom. This we have already seen in the proof of global consistency. Secondly, the soundness axioms of TLR (see below) are not completely interpretable. They include a stronger consequence for the order relation which does not hold in FSU. However, this property is, again, not essential for the applications.

TLR consists of four groups of axioms for the truth predicates. The *local truth axioms* are covered by the axioms II., III., and V. of FSU. The *level axioms* corresponds to the properties of our ordering relation, as proven in lemma 15, expect directedness. Of course, it would be follow in the presence of (U-Dir). But note that the unboundedness of the levels which follows in TLR from directedness and which is indeed needed follows in FSU already from the limit axiom. As *connection axioms* there are *limit*, *persistence*, *localization*, *potential completeness*, and *positive* and *negative soundness*. The interpretation of the limit axiom of TLR in FSU is lemma 12. Persistence was proved in lemma 15.1. Localization is trivial since we have $\dot{\mathsf{T}} \equiv \lambda x.x$. Potential completeness follows in FSU from axiom I.(1) together with the definition of the order relation. Positive soundness reads as $\mathsf{T}_j(\mathsf{T}_i\, x) \to i \preceq j \wedge \mathsf{T}_i(x)$. Here, the basic axiom I.(2) of FSU allows to derive from the premise the second conjunction. With respect to the order, we do not demand in FSU that $(t \mathrel{\dot\in} u) \in v$ already implies $u \sqsubset v$. Nevertheless, that it would come for free in our model construction, it looks like that it is not necessary for applications of FSU. Analogously, we can treat the essential part of negative soundness. Finally, TLR contains a *Reflection principle* which actually gives it its full strength:

$$\forall i.\forall y.\forall z.(\forall x.\exists j.\mathsf{T}_i(y\, x) \to \mathsf{T}_j(z\, x)) \to \exists k.\forall x.\exists j.j \preceq k \wedge (\mathsf{T}_i(y\, x) \to \mathsf{T}_j(z\, x))$$

According to the (informal) translation we would have as premise in FSU:

$$\mathsf{U}(u_i) \wedge \forall x.\exists u_j.\mathsf{U}(u_j) \wedge (\mathsf{T}(u_i\,(y\, x)) \to \mathsf{T}(u_j\,(z\, x))).$$

Considering the function $f := \lambda x.\dot\neg\,(u_i\,(y\, x))\mathrel{\dot\vee}(z\, x)$ it follows that f is a class. So limit yields a universe $\ell\, f$ reflecting f. From the closure condition of universes and T we get:

$$\begin{array}{rl} & \mathsf{T}(u_i\,(y\, x)) \to \mathsf{T}(u_j\,(z\, x)) \\ \to & \mathsf{T}(u_i\,(y\, x)) \to \mathsf{T}(z\, x) \\ \to & \mathsf{T}(\dot\neg\,(u_i\,(y\, x))\mathrel{\dot\vee}(z\, x)) \\ \to & (\dot\neg\,(u_i\,(y\, x))\mathrel{\dot\vee}(z\, x)) \in \ell\, f \\ \to & \dot\neg\,(u_i\,(y\, x)) \in \ell\, f \vee (z\, x) \in \ell\, f \\ \to & \mathsf{T}(\dot\neg\,(u_i\,(y\, x))) \vee (z\, x) \in \ell\, f \\ \to & \neg\mathsf{T}(u_i\,(y\, x)) \vee (z\, x) \in \ell\, f \\ \to & \mathsf{T}(u_i\,(y\, x)) \to \mathsf{T}(\ell\, f\,(z\, x)) \end{array}$$

Since $\ell\, f$ does not depend on x we get the antecedence by choosing $\ell\, f$ as the universe associated with k (and even the js) in the succedence of the reflection principle.

However, there is an important difference between FSU and TLR. In the former theory, the limit axiom is *uniform*. This will allow the definition of transfinite hierarchies of universes in FSU, cf. below.

Even without a formal embedding of TLR in FSU it is just a matter of routine to check that we can prove the features of TLR analogously in FSU. One has to inspect only whether directedness or the full consequence of the soundness axiom is used. This is not the case, at least, for the following applications which shows some of the main results about TLR and FSU.

First, the embedding of *explicit mathematics* into FON can be extended to "its" concept of *universes* as it was studied e.g. in [Fef82, Mar93, Mar94, Kah97b, MS98], cf. [Can96, §38]. Strahm gave in [Str99] a proof-theoretic analysis of the theories $\mathsf{EMU}\!\upharpoonright$, $\mathsf{EMU}\!\upharpoonright + (\Sigma^{+}\text{-}\mathsf{I_N})$, and EMU which matches in strength exactly with $\mathsf{FSU} + (\text{C-}\mathsf{I_N})$, $\mathsf{FSU} + (\text{T-}\mathsf{I_N})$, and $\mathsf{FSU} + (\mathcal{L}_U\text{-}\mathsf{I_N})$. But, in contrast to the theories of universes in explicit mathematics, FSU allows a syntactical embedding of *(transfinitely) iterated fixed point theories* $\widehat{\mathsf{ID}}_\alpha$, in analogy to the embedding of $\widehat{\mathsf{ID}}_1$ in $\mathsf{FON} + (\mathcal{L}_F\text{-}\mathsf{I_N})$.

Secondly, there is an embedding of the theory ATR_0 in $\mathsf{FSU} + (\text{C-}\mathsf{I_N})$, cf. [Can96, §40]. ATR_0, *arithmetical transfinite recursion*, is one of the main theories used in the program of *reverse mathematics*, cf. [Sim99]. Thus, FSU allows to formalize an important part of mathematics as it is shown for ATR_0, cf. [Sim99, chapter V].

Finally, truth theories allow a better inspection of set- and truth-theoretical paradoxes, in contrast to the theories of explicit mathematics. For instance, we can formulate the Russell "set" $\{x|x \notin x\}$. According to our abbreviations it reads as the term $r := \lambda x.x \dot{\notin} x = \lambda x.\dot{\neg}(x\,x)$. As for the "liar" used in the proof of lemma 2, we get $\mathsf{FON} \vdash \neg\mathsf{P}(r\,r)$. Now, we consider in addition the Russell set relativized to a universe: Let r be defined as $\lambda u, x.\dot{\neg}(u\,(x\,x))$. Thus, we have $\mathsf{r}\,u = \{x|(x \dot{\in} x) \notin u\}$. If we write r_u for $\mathsf{r}\,u$ we get $\mathsf{r}_u\,\mathsf{r}_u = \dot{\neg}(u\,(\mathsf{r}_u\,\mathsf{r}_u))$. As for FON we get that $\mathsf{r}_u\,\mathsf{r}_u$ can not be a class relative to a universe u:

Lemma 16 $\mathsf{FSU} \vdash \mathsf{U}(u) \rightarrow \neg(\mathsf{r}_u\,\mathsf{r}_u \in u \vee \dot{\neg}(\mathsf{r}_u\,\mathsf{r}_u) \in u)$

However, in a universe v which reflects u the term $\mathsf{r}_u\,\mathsf{r}_u$ becomes true, cf. also [Can96, lemma 37.5]:

Proposition 17 $\mathsf{FSU} \vdash \mathsf{U}(u) \wedge \mathsf{U}(v) \wedge u \sqsubset v \rightarrow \mathsf{r}_u\,\mathsf{r}_u \in v$

Proof: From the previous lemma we get $\neg(\mathsf{r}_u\,\mathsf{r}_u \in u)$. Since u is a universe we have $\mathsf{r}_u\,\mathsf{r}_u \notin u$. Thus, the definition of $\sqsubset$ yields $(\mathsf{r}_u\,\mathsf{r}_u \dot{\notin} u) \in v$. From the definition of r_u we get finally $\mathsf{r}_u\,\mathsf{r}_u \in v$. □

For the "liar" let us consider two different relativations: l is defined as $\mathsf{rec}\,(\lambda x.\dot{\neg}(u\,x))$. So we have $\mathsf{l}_u := \mathsf{l}\,u = \dot{\neg}(u\,\mathsf{l}_u)$. It can be read as "I'm false in u". The alternative is to define $\mathsf{m} := \mathsf{rec}\,(\lambda x.u\,(\dot{\neg}\,x))$. Here, we have $\mathsf{m}_u := \mathsf{m}\,u = u\,(\dot{\neg}\,\mathsf{m}_u)$ which can read as "My negation is true in u". Both cannot be propositions relativized to a univese u, but while l_u becomes true in universes reflecting u, for m_u it is its negation.

Proposition 18

1. $\mathsf{FSU} \vdash \mathsf{U}(u) \wedge \mathsf{U}(v) \wedge v \sqsubset u \rightarrow \neg(\mathsf{l}_u \in u \vee \dot{\neg}\,\mathsf{l}_u \in u) \wedge \mathsf{l}_u \in v$
2. $\mathsf{FSU} \vdash \mathsf{U}(u) \wedge \mathsf{U}(v) \wedge v \sqsubset u \rightarrow \neg(\mathsf{m}_u \in u \vee \dot{\neg}\,\mathsf{m}_u \in u) \wedge \dot{\neg}\,\mathsf{m}_u \in v$

The proof is a straightforward calculation, similar to the relativized Russell set.

Hierarchies of universes

In the following, we will investigate up to which length, depending on the induction principle, we can define hierarchies of universes in our theory.

Definition 19 A theory $\mathcal{T}$ proves *the existence of a hierarchy of universes of length* α, if there is term t such that:

$$\mathcal{T} \vdash \forall\beta \prec \alpha.\mathsf{U}(t\,\beta) \wedge \forall\gamma \prec \beta.t\,\gamma \sqsubset t\,\beta.$$

We say that $\mathcal{T}$ proves *the existence of hierarchies of universes of length* $< \alpha$, if it proves the existence of hierarchies of universes of length β for every β less than α.

By straightforward adaptation of the corresponding proofs in arithmetic [Sch77, Can89, JS96], we get the following results about transfinite induction in the presence of truth and formulae induction (see also [Can96, §46] for corresponding proofs over FON).

Proposition 20

1. For ordinals α less than ω^ω, we have:

$$\mathsf{FSU} + (\mathsf{T}\text{-}\mathrm{I}_{\mathsf{N}}) \vdash \mathrm{TI}(\alpha, \mathsf{T}(f\,\cdot)).$$

2. For ordinals α less than ε_0 and each $\mathcal{L}_U$ formula φ, it holds:

$$\mathsf{FSU} + (\mathcal{L}_U\text{-}\mathrm{I}_{\mathsf{N}}) \vdash \mathrm{TI}(\alpha, \varphi).$$

By use of transfinite induction, we will show that $\mathsf{FSU}+(\mathsf{T}\text{-}\mathrm{I}_{\mathsf{N}})$ proves the existence of hierarchies of universes of length $< \omega^\omega$, while $\mathsf{FSU} + (\mathcal{L}_U\text{-}\mathrm{I}_{\mathsf{N}})$ proves the existence of hierarchies of universes of length $< \varepsilon_0$. For this aim, we define an appropriate operation univ. By using recursion on the ordinals less then ε_0, we can define a term univ satisfying the following conditions:

1. $\mathsf{univ}\,0 = \ell\,(\lambda y.0 \dot{=} 0)$,
2. $\mathsf{univ}\,(\alpha + 1) = \ell\,(\mathsf{univ}\,\alpha)$,
3. $\mathsf{univ}\,\Lambda = \ell\,(\lambda y.\overbrace{\exists x \prec \Lambda.y \in \mathsf{univ}\,x}^{\cdot})$.

Lemma 21

1. $\mathsf{FSU} + \mathrm{TI}(\alpha, \mathsf{T}(f\,\cdot)) \vdash \forall\beta \prec \alpha.\mathsf{U}(\mathsf{univ}\,\beta)$,
2. $\mathsf{FSU} + \mathrm{TI}(\alpha, \mathsf{T}(f\,\cdot)) \vdash \forall\beta \prec \alpha.\forall\gamma \prec \beta.\mathsf{univ}\,\gamma \sqsubset \mathsf{univ}\,\beta$.

Proof: We prove both assertions by transfinite induction on β.

For the first claim, we have only to check that the arguments of the universe operator are classes. Note that the formula $\mathsf{C}(t)$ can be rewritten as the T-positive formula $\mathsf{T}(\dot{\forall} x.t\,x \mathrel{\dot{\vee}} \dot{\neg}\,(t\,x))$ such that transfinite induction for T formulae is sufficient.

The zero and successor cases are trivial. For the limit case, we have by the induction hypothesis $\forall x \prec \Lambda.\mathsf{U}(\mathsf{univ}\,x)$. So we get

$$\begin{array}{rcl} \forall x \prec \Lambda.\mathsf{U}(\mathsf{univ}\,x) & \rightarrow & \forall x \prec \Lambda.\mathsf{C}(\mathsf{univ}\,x) \\ & \rightarrow & \forall y.\forall x \prec \Lambda.y \in \mathsf{univ}\,x \vee y \notin \mathsf{univ}\,x \\ & \rightarrow & \forall y.(\exists x \prec \Lambda.y \in \mathsf{univ}\,x) \vee (\forall x \prec \Lambda.y \notin \mathsf{univ}\,x) \\ & \rightarrow & \forall y.\mathsf{T}(\overbrace{\exists x \prec \Lambda.y \in \mathsf{univ}\,x}^{\cdot}) \vee \mathsf{T}(\overbrace{\forall x \prec \Lambda.y \notin \mathsf{univ}\,x}^{\cdot}) \end{array}$$

$$\begin{aligned}
&\rightarrow \quad \forall y.\mathsf{T}(\overbrace{\exists x \prec \Lambda.y \in \mathsf{univ}\,x}^{\cdot}) \vee \mathsf{T}(\dot{\neg}(\overbrace{\exists x \prec \Lambda.y \in \mathsf{univ}\,x}^{\cdot}))\\
&\rightarrow \quad \mathsf{T}(\dot{\forall} z.(\lambda y.\overbrace{\exists x \prec \Lambda.y \in \mathsf{univ}\,x}^{\cdot})\,z \,\dot{\vee}\, \dot{\neg}((\lambda y.\overbrace{\exists x \prec \Lambda.y \in \mathsf{univ}\,x}^{\cdot})\,z))\\
&\rightarrow \quad \mathsf{C}(\lambda y.\overbrace{\exists x \prec \Lambda.y \in \mathsf{univ}\,x}^{\cdot})
\end{aligned}$$

Since $\mathsf{univ}\,\beta$ is a universe for all β less than α, we can prove the second assertion by case distinction on β. The zero and successor cases are obvious. The limit case follows from the closure conditions of universes. Let $\gamma \prec \Lambda$. Then we have:

$$\begin{aligned}
x \in \mathsf{univ}\,\gamma \quad &\rightarrow \quad (x \mathrel{\dot{\in}} \mathsf{univ}\,\gamma) \in \mathsf{univ}\,(\gamma+1)\\
&\rightarrow \quad \exists \delta \prec \Lambda.(x \mathrel{\dot{\in}} \mathsf{univ}\,\gamma) \in \mathsf{univ}\,\delta\\
&\rightarrow \quad \overbrace{\exists \delta \prec \Lambda.(x \mathrel{\dot{\in}} \mathsf{univ}\,\gamma) \in \mathsf{univ}\,\delta}^{\cdot} \in \mathsf{univ}\,\Lambda\\
&\rightarrow \quad \exists \delta \prec \Lambda.((x \mathrel{\dot{\in}} \mathsf{univ}\,\gamma) \mathrel{\dot{\in}} \mathsf{univ}\,\delta) \in \mathsf{univ}\,\Lambda\\
&\rightarrow \quad \exists \delta \prec \Lambda.(x \mathrel{\dot{\in}} \mathsf{univ}\,\gamma) \in \mathsf{univ}\,\Lambda\\
&\rightarrow \quad (x \mathrel{\dot{\in}} \mathsf{univ}\,\gamma) \in \mathsf{univ}\,\Lambda,
\end{aligned}$$

$$\begin{aligned}
x \notin \mathsf{univ}\,\gamma \quad &\rightarrow \quad (x \mathrel{\dot{\notin}} \mathsf{univ}\,\gamma) \in \mathsf{univ}\,(\gamma+1)\\
&\rightarrow \quad \exists \delta \prec \Lambda.(x \mathrel{\dot{\notin}} \mathsf{univ}\,\gamma) \in \mathsf{univ}\,\delta\\
&\rightarrow \quad \overbrace{\exists \delta \prec \Lambda.(x \mathrel{\dot{\notin}} \mathsf{univ}\,\gamma) \in \mathsf{univ}\,\delta}^{\cdot} \in \mathsf{univ}\,\Lambda\\
&\rightarrow \quad \exists \delta \prec \Lambda.((x \mathrel{\dot{\notin}} \mathsf{univ}\,\gamma) \mathrel{\dot{\in}} \mathsf{univ}\,\delta) \in \mathsf{univ}\,\Lambda\\
&\rightarrow \quad \exists \delta \prec \Lambda.(x \mathrel{\dot{\notin}} \mathsf{univ}\,\gamma) \in \mathsf{univ}\,\Lambda\\
&\rightarrow \quad (x \mathrel{\dot{\notin}} \mathsf{univ}\,\gamma) \in \mathsf{univ}\,\Lambda.
\end{aligned}$$

□

From proposition 20 and the trivial fact that $\mathrm{TI}(k, \mathsf{T}(f\,\cdot))$ is provable in $\mathsf{FON} + (\mathsf{C}\text{-}\mathrm{I}_{\mathsf{N}})$ for every $k < \omega$, we get:

Proposition 22

1. $\mathsf{FSU} + (\mathsf{C}\text{-}\mathrm{I}_{\mathsf{N}})$ proves the existence of hierarchies of universes of length $< \omega$.
2. $\mathsf{FSU} + (\mathsf{T}\text{-}\mathrm{I}_{\mathsf{N}})$ proves the existence of hierarchies of universes of length $< \omega^{\omega}$.
3. $\mathsf{FSU} + (\mathcal{L}_U\text{-}\mathrm{I}_{\mathsf{N}})$ proves the existence of hierarchies of universes of length $< \varepsilon_0$.

Fixed points

In the following, we show that hierarchies of universes allow us to define iterated fixed points. These fixed points will be used for the embedding of the the theories $\widehat{\mathsf{ID}}_{\alpha}$, which are defined in the next section, into Frege structures with universes. This fact shows the close relationship of universes and fixed points. We will call a $\mathcal{L}_U$ formula $\varphi(P, Q, x, y)$ an *inductive operator form* if it is T-positive and contains the relation variable P only positively (while Q is allowed to occur positively and negatively). To state the following proposition in a readable way, we need some abbreviations. We will use the usual short hand notation for pairing and projection, i.e. (t, s) for $(\mathsf{p}\,t\,s)$, $(t)_0$ for $(\mathsf{p}_0\,t)$, and $(t)_1$ for $(\mathsf{p}_1\,t)$. Let $t_{\beta}\,s$ be an abbreviation of $t\,(s, \beta)$ and $t_{\prec\beta}\,r$ of $((t\,r) \mathrel{\dot{\in}} \mathsf{univ}\,(r)_1) \mathrel{\dot{\wedge}} ((r)_1 \mathrel{\dot{\prec}} \beta)$. To deal with the negative occurrences of Q, sometimes we use the notation $\varphi(P, Q, \neg Q, x, y)$ where the second argument stands for the positive occurrences of Q only, while the third indicates the negative ones.

Proposition 23 Let $\varphi(P,Q,x,y)$ be an inductive operator form. Then there exists a term t^φ of $\mathcal{L}_U$ such that

$$\mathsf{FSU}+\mathrm{TI}(\alpha,\mathsf{T}(f\,\cdot))\vdash\forall\beta\prec\alpha.\forall x.\mathsf{T}(t^\varphi_\beta\,x)\leftrightarrow\varphi(\mathsf{T}(t^\varphi_\beta\,\cdot),\mathsf{T}(t^\varphi_{\prec\beta}\,\cdot),x,\beta).$$

Proof: Given a inductive operator form $\varphi(P,Q,x,y)$, we choose fresh variables u and z not occurring in $\varphi(P,Q,x,y)$. z is the variable for the fixed point term which will be defined by the recursion theorem, and u stands for its argument. Thus, it will often be decomposed in two components, $(u)_0$ for the "real" argument and $(u)_1$ for the level. Now we do the following replacements in the formula $\varphi(P,Q,\neg Q,x,y)$:

- the variable x is replaced by $(u)_0$,
- the variable y is replaced by $(u)_1$,
- all occurrences of $P(t)$ are replaced by $\mathsf{T}(z\,(t,(u)_1))$,
- the positive occurrences of $Q(s)$ are replaced by $z\,s\in\mathsf{univ}\,(s)_1\wedge(s)_1\prec(u)_1$,
- the negative occurrences $\neg Q(r)$ are replaced by $(z\,r\mathrel{\dot{\notin}}\mathsf{univ}\,(r)_1)\in\mathsf{univ}\,(u)_1\vee\neg((r)_1\prec(u)_1)$.

The last clause is crucial. Because of the reflection of the statement $z\,r\notin\mathsf{univ}\,(r)_1$ in the universe $\mathsf{univ}\,(u)_1$ the formula will become T-positive.

Thus, the whole formula abbreviated by $\varphi(z,u)$ reads

$$\varphi(\mathsf{T}(z\,(\cdot,(u)_1)),z\cdot\in\mathsf{univ}\,(\cdot)_1\wedge(\cdot)_1\prec(u)_1,(z\cdot\mathrel{\dot{\notin}}\mathsf{univ}\,(\cdot)_1)\in\mathsf{univ}\,(u)_1\vee\neg((\cdot)_1\prec(u)_1),(u)_0,(u)_1).$$

Now we define:

$$t^\varphi:=\mathsf{rec}\,(\lambda z,u.\overset{\cdot}{\overbrace{\varphi(z,u)}}).$$

So we have:

$$\begin{aligned}
&t^\varphi\,(x,\beta)\\
=\;&\overset{\cdot}{\overbrace{\varphi(\mathsf{T}(t^\varphi\,(\cdot,\beta)),t^\varphi\cdot\in\mathsf{univ}\,(\cdot)_1\wedge(\cdot)_1\prec\beta,(t^\varphi\cdot\mathrel{\dot{\notin}}\mathsf{univ}\,(\cdot)_1)\in\mathsf{univ}\,\beta\vee\neg((\cdot)_1\prec\beta),x,\beta)}}\\
=\;&\overset{\cdot}{\overbrace{\varphi(\mathsf{T}(t^\varphi_\beta\,\cdot),\mathsf{T}(t^\varphi_{\prec\beta}\,\cdot),(t^\varphi\cdot\mathrel{\dot{\notin}}\mathsf{univ}\,(\cdot)_1)\in\mathsf{univ}\,\beta\vee\neg((\cdot)_1\prec\beta),x,\beta)}}.
\end{aligned}$$

From the assumption $\beta\prec\alpha$, we get with lemma 21 that $\mathsf{U}(\mathsf{univ}\,\beta)$ holds. Now, $\forall x.\mathsf{T}(t^\varphi_\beta\,x)\leftrightarrow\varphi(\mathsf{T}(t^\varphi_\beta\,\cdot),\mathsf{T}(t^\varphi_{\prec\beta}\,\cdot),x,\beta)$ follows by straightforward induction on the build up of φ. The only cases to check are the negative occurrences of $\mathsf{T}(t^\varphi_{\prec\beta}\,\cdot)$. So we have to show the equivalence of $\mathsf{T}(\overset{\cdot}{\overbrace{(t^\varphi\,r\mathrel{\dot{\notin}}\mathsf{univ}\,(r)_1)\in\mathsf{univ}\,\beta\vee\neg((r)_1\prec\beta)}})$ and $\neg\mathsf{T}(t^\varphi_{\prec\beta}\,r)$:

$$\begin{aligned}
&\mathsf{T}(\overset{\cdot}{\overbrace{(t^\varphi\,r\mathrel{\dot{\notin}}\mathsf{univ}\,(r)_1)\in\mathsf{univ}\,\beta\vee\neg((r)_1\prec\beta)}})\\
\leftrightarrow\;&(t^\varphi\,r\mathrel{\dot{\notin}}\mathsf{univ}\,(r)_1)\in\mathsf{univ}\,\beta\vee\neg((r)_1\prec\beta)\\
\leftrightarrow\;&((r)_1\prec\beta)\rightarrow(t^\varphi\,r\mathrel{\dot{\notin}}\mathsf{univ}\,(r)_1)\in\mathsf{univ}\,\beta\\
\leftrightarrow\;&((r)_1\prec\beta)\rightarrow t^\varphi\,r\notin\mathsf{univ}\,(r)_1\\
\leftrightarrow\;&\neg(t^\varphi\,r\in\mathsf{univ}\,(r)_1)\vee\neg((r)_1\prec\beta)\\
\leftrightarrow\;&\neg(\mathsf{T}(t^\varphi\,r\mathrel{\dot{\in}}\mathsf{univ}\,(r)_1))\wedge\mathsf{T}(\overset{\cdot}{\overbrace{(r)_1\prec\beta}})\\
\leftrightarrow\;&\neg\mathsf{T}(\overset{\cdot}{\overbrace{(t^\varphi\,r\in\mathsf{univ}\,(r)_1)\wedge((r)_1\prec\beta)}})\\
\leftrightarrow\;&\neg\mathsf{T}(t^\varphi_{\prec\beta}\,r).
\end{aligned}$$

□

5 Proof-theoretic investigations

Embeddings of $\widehat{\mathsf{ID}}_\alpha$

To determine the proof theoretic lower bounds of the theories of universes over Frege structures, we will interpret *iterated fixed point theories* $\widehat{\mathsf{ID}}_\alpha$ into FSU plus the different induction principles. As for FON, we have to stress the fact that we not only have a proof-theoretic reduction, but a syntactical interpretation which shows the expressive power of our theories.

In the following, we need *transfinitely* iterated fixed point theories, not only finitely iterated ones. The latter ones are well-known from Feferman's proof of Hancock's conjecture [Fef82]. The transfinite case was recently analyzed by Jäger et. al. [JKSS99].

These theories $\widehat{\mathsf{ID}}_\alpha$ are formulated in the language $\mathcal{L}_{\mathsf{fix}}$, which expands the language $\mathcal{L}_1$ of Peano arithmetic by predicate constants $\mathcal{P}^\varphi$ for each inductive operator form $\varphi(P, Q, x, y)$, i.e. a $\mathcal{L}_1$ formula, containing $P(t)$ at most positively, while $Q(s)$ is allowed to occur positively and negatively. Here, P and Q have to be fresh (unary) relation variables.

If $\langle\cdot,\cdot\rangle$ is a primitive recursive pairing function with projections $[\cdot]_0$ and $[\cdot]_1$, we write $\mathcal{P}^\varphi_s(t)$ for $\mathcal{P}^\varphi(\langle t, s\rangle)$ and $\mathcal{P}^\varphi_{\prec s}(t)$ for $t = \langle [t]_0, [t]_1\rangle \wedge [t]_1 \prec s \wedge \mathcal{P}^\varphi(t)$. The parameter s in $\mathcal{P}^\varphi_s(t)$ is understood as the level of the fixed point definition. With $\mathcal{P}^\varphi_{\prec s}(t)$, it is expressed that t belongs to the disjoint union of fixed points with levels less than s.

The axioms of $\widehat{\mathsf{ID}}_\alpha$, for α an ordinal less than ε_0, are those of PA together with induction on the natural numbers for all $\mathcal{L}_{\mathsf{fix}}$ formulae plus the following fixed point axioms for each inductive operator form $\varphi(P, Q, x, y)$:

$$\forall \beta \prec \alpha. \forall x. \mathcal{P}^\varphi_\beta(x) \leftrightarrow \varphi(\mathcal{P}^\varphi_\beta, \mathcal{P}^\varphi_{\prec\beta}, x, \beta).$$

$\widehat{\mathsf{ID}}_{<\alpha}$ is the union of the theories $\widehat{\mathsf{ID}}_\beta$, $\beta < \alpha \le \varepsilon_0$.

The theories $\widehat{\mathsf{ID}}_\alpha$ are the metapredicative counterparts to the well-known impredicative theories of inductive definitions ID_α which contains, additionally, the leastness of the fixed points.

The proof-theoretic analysis of ID_α in [JKSS99] yields the following theorem about the proof-theoretic strength. Here, $\boldsymbol{\varphi}$ denotes the ternary Veblen function which is the generalization of the well-known binary Veblen function, see above. In particular, $\boldsymbol{\varphi}100$ is the Feferman-Schütte ordinal Γ_0.

Theorem 24

1. $|\widehat{\mathsf{ID}}_{<\omega}| = \boldsymbol{\varphi}100 = \Gamma_0$
2. $|\widehat{\mathsf{ID}}_{<\omega^\omega}| = \boldsymbol{\varphi}1\omega0$
3. $|\widehat{\mathsf{ID}}_{<\varepsilon_0}| = \boldsymbol{\varphi}1\varepsilon_0 0$

To embed the $\widehat{\mathsf{ID}}_\alpha$ theories into FSU, we extend the standard interpretation $\cdot^N$ of PA in TON by setting

$$(\mathcal{P}^\varphi_\beta(s))^N \quad := \quad \mathsf{T}(t^{(\varphi^N)}_\beta(s^N))$$

where $t^{(\varphi^N)}_\beta$ is given as in proposition 23. Using this proposition, the verification of the fixed point axiom of $\widehat{\mathsf{ID}}_\alpha$ in $\mathsf{FSU} + \mathrm{TI}(\alpha, \mathsf{T}(f\,\cdot))$ is obvious. Of course, the induction principle of $\widehat{\mathsf{ID}}_\alpha$ can be verified by ($\mathcal{L}_U$-I_{N}). Moreover, in $\mathsf{FSU} + \mathrm{TI}(\alpha+1, \mathsf{T}(f\,\cdot))$ we can use (C-I_{N}) for this verification, since the (translations of) formulae containing (possibly negative) occurrences of fixed points up to level α can be rewritten as classes by use of the universe $\mathsf{univ}\,(\alpha+1)$. So we have:

Proposition 25 For each ordinal α less than ε_0, it holds that:

1. $\widehat{\mathsf{ID}}_\alpha \vdash \varphi \quad \Rightarrow \quad \mathsf{FSU} + \mathrm{TI}(\alpha, \mathsf{T}(f\,\cdot)) + (\mathcal{L}_U\text{-}\mathrm{I}_\mathsf{N}) \vdash \varphi^N$,
2. $\widehat{\mathsf{ID}}_\alpha \vdash \varphi \quad \Rightarrow \quad \mathsf{FSU} + \mathrm{TI}(\alpha+1, \mathsf{T}(f\,\cdot)) + (\mathsf{C}\text{-}\mathrm{I}_\mathsf{N}) \vdash \varphi^N$.

Therefore we get, together with proposition 22, the following embeddings for the limit cases:

Proposition 26 There exists a translation ^{N} from the language $\mathcal{L}_{\mathsf{fix}}$ in the language $\mathcal{L}_U$ such that

1. $\widehat{\mathsf{ID}}_{<\omega} \vdash \varphi \quad \Rightarrow \quad \mathsf{FSU} + (\mathsf{C}\text{-}\mathrm{I}_\mathsf{N}) \vdash \varphi^N$,
2. $\widehat{\mathsf{ID}}_{<\omega^\omega} \vdash \varphi \quad \Rightarrow \quad \mathsf{FSU} + (\mathsf{T}\text{-}\mathrm{I}_\mathsf{N}) \vdash \varphi^N$,
3. $\widehat{\mathsf{ID}}_{<\varepsilon_0} \vdash \varphi \quad \Rightarrow \quad \mathsf{FSU} + (\mathcal{L}_U\text{-}\mathrm{I}_\mathsf{N}) \vdash \varphi^N$.

The upper bounds

The proof of the upper bounds follows a well-known procedure. First we define a Tait calculus for FSU which allows us to prove partial cut elimination up to $\mathsf{T}^+/\mathsf{T}^-$ formulae (see definition 27 below). Then, we introduce partial models of FSU by use of an appropriate inductive operator form which can be formalized in $\widehat{\mathsf{ID}}_\alpha$. As for FON we have to extend these theories slightly to deal with consistency. The partial models can be used to give an asymmetric interpretation of our theories. This procedure is well-known from the literature, cf. e.g. [Sch77, Jäg80, Jäg84, Can85, Can86] and especially Cantini [Can89, Can96] for truth theories or Glaß, Marzetta and Strahm [Mar93, Mar94, Gla93, Gla95, GS96, MS98, Str99] for explicit mathematics.

A Tait calculus for FSU

For the proof-theoretic treatment, we replace our Hilbert-style calculus by a Tait-style calculus, [Tai68]. Thus, we dispense of logic with negation, but start with the connectives $\wedge, \vee$ and quantifiers $\forall$ and $\exists$. Additionally, we extend the language $\mathcal{L}_U$ with new relation symbols

- $\widetilde{\mathsf{N}}, \widetilde{=}, \widetilde{\mathsf{T}}$ and $\widetilde{\mathsf{U}}$

for the negation of the corresponding relation symbols of $\mathcal{L}_U$. In the usual way, we define negation for arbitrary formulae of the resulting language $\mathcal{L}_U^T$. We deal with finite sets of $\mathcal{L}_U^T$ formula $(\Gamma, \Delta, \ldots)$. The intended logical meaning of such a set is the disjunction of its elements. We will use the usual short hand notation, e.g. φ is used for the singleton $\{\varphi\}$, and we abbreviate union of sets by use of commas.

Definition 27 A $\mathcal{L}_U^T$ formula φ is called a T^+ formula, if it contains no subformulas of the form $\widetilde{\mathsf{T}}(t)$ and $\widetilde{\mathsf{U}}(s)$; it is called T^- formula, if it contains none of the form $\mathsf{T}(t)$ and $\mathsf{U}(s)$.

To deal with class and truth induction, we introduce the sequent calculus FSU^T. For formulae induction, we have to define a semi-formal system FSU^∞ which will be described later.

The logical axioms and rules of FSU^T are as usual, containing equality and the cut rule:

$$(\text{Cut}) \quad \frac{\Gamma, \varphi \quad \Gamma, \neg\varphi}{\Gamma}$$

where φ is called the cut formula of this inference.

The non-logical axioms are as usual. The axiom of TON can be chosen in the standard way. To ensure that all main formulae are T^+ or T^- formulae, we have to split the closure conditions of FSU into two axioms. For instance, the closure under double negation of FSU reads as

(7a) $\quad \Gamma, \widetilde{\mathsf{U}}(s), \widetilde{\mathsf{T}}(s\,t), \mathsf{T}(s\,(\dot\neg\,(\dot\neg\,t)))$, (7b) $\Gamma, \widetilde{\mathsf{U}}(s), \widetilde{\mathsf{T}}(s\,(\dot\neg\,(\dot\neg\,t))), \mathsf{T}(s\,t)$.

In addition, we get for the basic axioms, order structure, local consistency, and the limit axiom of FSU the following axioms in FSU^T:

(1) $\Gamma, \widetilde{\mathsf{U}}(s), \mathsf{T}(s\,t) \vee \mathsf{T}(\dot\neg\,(s\,t))$,

(2) $\Gamma, \widetilde{\mathsf{U}}(s), \widetilde{\mathsf{T}}(s\,t), \mathsf{T}(t)$,

(12) $\Gamma, \widetilde{\mathsf{U}}(r), \widetilde{\mathsf{U}}(s), \widetilde{\mathsf{T}}(s\,(r\,t)), \mathsf{T}(s\,t)$,

(13) $\Gamma, \widetilde{\mathsf{U}}(s), \widetilde{\mathsf{T}}(s\,t), \widetilde{\mathsf{T}}(s\,(\dot\neg\,t))$,

(14a) $\Gamma, (\exists x.\widetilde{\mathsf{T}}(t\,x) \wedge \widetilde{\mathsf{T}}(\dot\neg\,(t\,x))), \mathsf{U}(\ell\,t)$,

(14b) $\Gamma, (\exists x.\widetilde{\mathsf{T}}(t\,x) \wedge \widetilde{\mathsf{T}}(\dot\neg\,(t\,x))), \widetilde{\mathsf{T}}(t\,s), \mathsf{T}(\ell\,t\,(t\,s))$,

(14c) $\Gamma, (\exists x.\widetilde{\mathsf{T}}(t\,x) \wedge \widetilde{\mathsf{T}}(\dot\neg\,(t\,x))), \widetilde{\mathsf{T}}(\dot\neg\,(t\,s)), \mathsf{T}(\ell\,t\,(\dot\neg\,(t\,s)))$.

For class and truth induction, we have to use in FSU^T the following rules:

$$(\mathsf{C}\text{-}\mathsf{I}_\mathsf{N})^T \quad \frac{\Gamma, \mathsf{C}(t) \quad \Gamma, \mathsf{T}(t\,0) \quad \Gamma, \forall x : \mathsf{N}.\mathsf{T}(\dot\neg\,(t\,x)) \vee \mathsf{T}(t\,(\mathsf{s_N}\,x))}{\Gamma, \forall x : \mathsf{N}.\mathsf{T}(t\,x)}$$

where the assumption $\mathsf{C}(t)$ allows us to write the induction step as a T-positive formula, and

$$(\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})^T \quad \frac{\Gamma, \mathsf{T}(t\,0) \quad \Gamma, \forall x : \mathsf{N}.\mathsf{T}(t\,x) \rightarrow \mathsf{T}(t\,(\mathsf{s_N}\,x))}{\Gamma, \forall x : \mathsf{N}.\mathsf{T}(t\,x)}$$

The main formula of both rules is $\forall x : \mathsf{N}.\mathsf{T}(t\,x)$.

The semi-formal system FSU^∞ is defined as usual. (For detailed definitions of semi-formal systems in the context of explicit mathematics, cf. [GS96, MS98], or for truth theories [Can96].) We can take over all axiom and rules from FSU^T in the formalism of FSU^∞, adding the following ω rule to deal with formula induction (here, $\underline{n}$ denotes the nth numeral):

$$(\omega) \quad \frac{\Gamma, \underline{n} \,\widetilde{=}\, t \quad \text{for all } n < \omega}{\Gamma, \widetilde{\mathsf{N}}(t)}$$

Let φ^T be the canonical translation of a $\mathcal{L}_U$ formula φ in the Tait calculus. Then, we have:

Proposition 28

1. $\mathsf{FSU} + (\mathsf{C}\text{-}\mathsf{I}_\mathsf{N}) \vdash \varphi \quad \Rightarrow \quad \mathsf{FSU}^T + (\mathsf{C}\text{-}\mathsf{I}_\mathsf{N})^T \models \varphi^T$,
2. $\mathsf{FSU} + (\mathsf{T}\text{-}\mathsf{I}_\mathsf{N}) \vdash \varphi \quad \Rightarrow \quad \mathsf{FSU}^T + (\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})^T \models \varphi^T$,
3. $\mathsf{FSU} + (\mathcal{L}_U\text{-}\mathsf{I}_\mathsf{N}) \vdash \varphi \quad \Rightarrow \quad \mathsf{FSU}^\infty \models \varphi^T$.

We say that a derivation of Γ is *quasi-normal* if each cut formula in the derivation is a T^+ or T^- formula. For a quasi-normal derivation of Γ with length $k < \omega$ in FSU^T or $\alpha < \varepsilon_0$ in FSU^∞ we write $\models^{k}_{\star} \Gamma$ (or $\models^{<\omega}_{\star} \Gamma$) and $\models^{\alpha}_{\star} \Gamma$ (or $\models^{<\varepsilon_0}_{\star} \Gamma$), respectively.

Because all main formulae of the non-logical axioms and rules of FSU^T and FSU^∞ are T^+ or T^- formulae, we can prove partial cut elimination by use of the usual methods of proof theory, cf. e.g. [Sch77] or [Mar94].

Proposition 29 For every finite set Γ of $\mathcal{L}_U^T$ formulae, it holds that:

1. $\mathsf{FSU}^T + (\mathsf{C}\text{-}\mathsf{I}_\mathsf{N})^T \vdash \Gamma \quad \Rightarrow \quad \mathsf{FSU}^T + (\mathsf{C}\text{-}\mathsf{I}_\mathsf{N})^T \models^{<\omega}_{\star} \Gamma$,
2. $\mathsf{FSU}^T + (\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})^T \vdash \Gamma \quad \Rightarrow \quad \mathsf{FSU}^T + (\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})^T \models^{<\omega}_{\star} \Gamma$,
3. $\mathsf{FSU}^\infty \vdash \Gamma \quad \Rightarrow \quad \mathsf{FSU}^\infty \models^{<\varepsilon_0}_{\star} \Gamma$.

Partial models

The following model construction is an adaptation of Strahm's model for the theory EMU, explicit mathematics with universes, in [Str99].

We start with a given formalization of $\mathcal{CTT}$, the standard model of TON, in PA. Let $\mathcal{N}$ be the interpretation of N.

First, we extend this formalization to the additional constants $\dot{=}, \dot{\mathsf{N}}, \dot{\neg}, \dot{\wedge}, \dot{\forall}$ and ℓ of $\mathcal{L}_U$. Since there are no new equalities about these constants, the only requirement is the uniqueness of the formalized terms, e.g. $t \dot{=} s$ has to be different from $t \dot{\wedge} s$. To keep the notation simple, in the following, we do not distinguish between the terms of $\mathcal{L}_U$ and its formalization in the model.

Now we define an inductive operator form $\Phi(X, Y, x, \alpha)$ which captures the closure conditions of the predicate T.

$\Phi(X, Y, a, \alpha)$ holds if there exist elements b, c and f of the model such that one of the following clauses holds. (Here $\mathcal{S}(\alpha)$ means that α is a successor ordinal.)

1. $a \in Y$,
2. $\alpha = 0 \wedge a = (b \dot{=} c) \wedge b = c$,
3. $\alpha = 0 \wedge a = \dot{\neg}\,(b \dot{=} c) \wedge b \neq c$,
4. $\alpha = 0 \wedge a = (\dot{\mathsf{N}}\, b) \wedge \mathcal{N}(b)$,
5. $\alpha = 0 \wedge a = \dot{\neg}(\dot{\mathsf{N}}\, b) \wedge \neg\mathcal{N}(b)$,
6. $a = \dot{\neg}(\dot{\neg}\, b) \wedge a \notin Y \wedge b \in X$,
7. $a = (b \,\dot{\wedge}\, c) \wedge a \notin Y \wedge b \in X \wedge c \in X$,
8. $a = \dot{\neg}(b \,\dot{\wedge}\, c) \wedge a \notin Y \wedge ((\dot{\neg}\, b \in X) \vee (\dot{\neg}\, c \in X))$,
9. $a = (\dot{\forall}\, f) \wedge a \notin Y \wedge \forall x. f\, x \in X$,
10. $a = \dot{\neg}\,(\dot{\forall}\, f) \wedge a \notin Y \wedge \exists x. \dot{\neg}\,(f\, x) \in X$,
11. $\mathcal{S}(\alpha) \wedge a = (\ell\, b)\, c \wedge a \notin Y \wedge \dot{\neg}\, a \notin Y \wedge (\forall x. b\, x \in Y \vee \dot{\neg}\,(b\, x) \in Y) \wedge c \in Y$,
12. $\mathcal{S}(\alpha) \wedge a = \dot{\neg}\,((\ell\, b)\, c) \wedge a \notin Y \wedge \dot{\neg}\, a \notin Y \wedge (\forall x. b\, x \in Y \vee \dot{\neg}\,(b\, x) \in Y) \wedge c \notin Y$.

In the proof-theoretic analysis of FSU we have to consider not only fixed points of Φ but *consistent fixed points* which are needed to verify the consistency axiom. But before we turn to these considerations let us describe briefly how we can build a full model of $\mathsf{FSU} + (\mathcal{L}_U\text{-}\mathrm{I}_{\mathsf{N}})$ by use of Φ. Let us define a Φ-sequence (Z_β) of *least fixed points* Z_β for $\beta \leq \alpha$ in the following way:

1. Z_0 is the empty set $\emptyset$.
2. $Z_{\beta+1}$ is the least fixed point of $\Phi(Z, Z_\beta, x, \beta + 1)$,
3. If β is a limit ordinal, then Z_β is the least fixed point of $\Phi(X, \bigcup_{\gamma<\beta} X_\gamma, x, \beta)$.

Iterating this sequence up to the *first recursively inaccessible ordinal* ι we get a model $\mathcal{I}_\iota$ of $\mathsf{FSU} + (\mathcal{L}_U\text{-}\mathrm{I}_{\mathsf{N}})$. (For the notion of recursively inaccessible ordinal, and admissible ordinal used below, we refer to the standard literature, e.g. [Bar75, Hin78].) The construction and the proof is analogous to the corresponding model $\mathcal{C}_\iota$ of TLR by Cantini, cf. [Can96, theorem 39.16]. $\mathsf{T}(t)$ can be interpreted in $\mathcal{I}_\iota$ by $\exists\gamma < \iota. t \in Z_\gamma$, and $\mathsf{U}(t)$ by $\exists\beta < \iota. \exists b. t = \ell\, b \wedge \forall x. b\, x \in Z_\beta \vee \dot{\neg}\,(b\, x) \in Z_\beta$. In the verification that $\mathcal{I}_\iota$ builds a model consistency follows directly from the leastness of the fixed points. The only complicated case is the limit axioms. For it we get from the premise

$$\mathcal{I}_\iota \models \forall x. \exists\gamma. f\, x \in Z_\gamma \vee \dot{\neg}\,(f\, x) \in Z_\gamma.$$

As in the proof of theorem 39.16 of [Can96] one can check that this formula can be rewritten by use of a Σ_1 formula and Σ_1-reflection available in L_ι, the constructible hierarchy up to ι, allows to conclude

$$\mathcal{I}_\iota \models \exists\delta\forall x.f\,x \in Z_\delta \vee \dot\neg\,(f\,x) \in Z_\delta.$$

Now, one can easily check that $\ell\,f$ will be a universe in $Z_{\delta+1}$ "reflecting" f. To find δ ι has to be at least *admissible*, and because this can be iterated, ι has to be actually a limit of *admissibles*, i.e. it has to be *recursively inaccessible*.

From the definition of Φ it is already clear, that $\mathcal{I}_\iota$ will also satisfy the axiom of linearity of universes (U-Lin), and therefore also (U-Dir), as well as (U-Nor). In fact, we could even demand that the hierarchy of universes is *well-founded*. However while the former axioms will not change the proof-theoretic strength, this is not obvious for an axiom stating the well-foundedness of the universes.

To determine the proof-theoretic upper bounds of FSU we have to work with much shorter sequences. But, this time we have to build in consistency in the fixed points. A set X is called a *consistent fixed point* of Φ, if it is a fixed point of Φ and, in addition, for all elements a it holds $\neg(a \in X \wedge \dot\neg\,a \in X)$. We call a sequence of sets (X_β) for $\beta \leq \alpha$ a Φ^+ sequence if the following conditions hold:

1. X_0 is a consistent fixed point of $\Phi(X, \emptyset, x, 0)$,
2. $X_{\beta+1}$ is a consistent fixed point of $\Phi(X, X_\beta, x, \beta+1)$,
3. If β is a limit ordinal, then X_β is a consistent fixed point of $\Phi(X, \bigcup_{\gamma<\beta} X_\gamma, x, \beta)$.

Now we define partial models $\mathcal{M}(\alpha)$ of FSU by use of a Φ^+ sequence $(X_\beta)_{\beta\leq\alpha}$. Given the interpretation of TON and the extension to the new constants, we have to interpret the predicates T and U. We translate the truth predicate $\mathsf{T}(t)$ by

$$t \in X_\alpha$$

and the universe predicate $\mathsf{U}(t)$ by

$$\exists\beta < \alpha.\exists b.t = \ell\,b \wedge \forall x.b\,x \in X_\beta \vee \dot\neg\,(b\,x) \in X_\beta.$$

The negative relation symbols $\widetilde{\mathsf{T}}(t)$ and $\widetilde{\mathsf{U}}(t)$ are interpreted by the corresponding negations. Let us introduce the following abbreviations:

- $\mathsf{T}_\beta(t)$ for $t \in X_\beta$, $\beta \leq \alpha$.
- $\mathsf{C}_\beta(t)$ for $\forall x.t\,x \in X_\beta \vee \dot\neg\,(t\,x) \in X_\beta$, i.e. t is a class at level β.
- $\mathsf{U}_\gamma(t)$ for $\exists b.t = \ell\,b \wedge \mathsf{C}_\beta(b)$, if γ is the successor ordinal $\beta + 1$.

Thus, the translation of $\mathsf{U}(t)$ reads as $\exists\beta < \alpha.\mathsf{U}_{\beta+1}(t)$. Please not, that the condition $\mathcal{S}(\alpha)$ in the clauses 11. and 12. of the definition of Φ in fact guarantees that for any term t the minimal γ for which $\mathsf{U}_\gamma(t)$ can hold has to be a successor ordinal.

To give an asymmetric interpretation of FSU^T into the structures $\mathcal{M}(\alpha)$, the following persistence properties are crucial. The proof follows immediately from the definition of $\mathcal{M}(\alpha)$. With the notation $\Gamma_{\vec{x}}$ or $\varphi_{\vec{x}}$, we indicate that all free variables of Γ or φ belong to the list $\vec{x}$, and $\Gamma_{\vec{n}}$ and $\varphi_{\vec{n}}$ stand for the corresponding substitution of the elements of $\vec{n}$ for $\vec{x}$.

Lemma 30 Let $\mathcal{M}(\alpha)$ be a structure for $\mathcal{L}_U$ and let $\gamma \leq \beta \leq \alpha$. Then, we have for all T^+ formulae $\varphi_{\vec{x}}$, all T^- formulae $\psi_{\vec{x}}$ and all elements $\vec{n}$ of $\mathcal{M}(\alpha)$:

1. $\mathcal{M}(\gamma) \models \varphi_{\vec{n}} \quad \Rightarrow \quad \mathcal{M}(\beta) \models \varphi_{\vec{n}}$,
2. $\mathcal{M}(\beta) \models \psi_{\vec{n}} \quad \Rightarrow \quad \mathcal{M}(\gamma) \models \psi_{\vec{n}}$.

Given as set $\Gamma_{\vec{x}}$ of T^+ and T^- formulae, we define for a given structure $\mathcal{M}(\alpha)$ and $\gamma \leq \beta \leq \alpha$ the expression

$$\mathcal{M}(\gamma, \beta) \models \Gamma_{\vec{n}}$$

for the fact that there is a T^- formula $\psi_{\vec{x}}$ in Γ such that

$$\mathcal{M}(\gamma) \models \psi_{\vec{n}}$$

or there is T^+ formula $\varphi_{\vec{x}}$ in Γ such that

$$\mathcal{M}(\beta) \models \varphi_{\vec{n}}.$$

Proposition 31 Let $\Gamma_{\vec{x}}$ be a finite set of T^+ and T^- formulae. Then we have for natural numbers k and ordinals $\alpha < \varepsilon_0$:

1. $\mathsf{FSU}^T + (\mathsf{C\text{-}I_N})^T \mid\!\frac{k}{\star}\; \Gamma_{\vec{x}} \quad \Rightarrow \quad$ For all $m > 0$ we have: $\mathcal{M}(m, m + 2^k) \models \Gamma_{\vec{n}}$,
2. $\mathsf{FSU}^T + (\mathsf{T\text{-}I_N})^T \mid\!\frac{k}{\star}\; \Gamma_{\vec{x}} \quad \Rightarrow \quad$ For all $\beta > 0$ we have: $\mathcal{M}(\beta, \beta + \omega^k) \models \Gamma_{\vec{n}}$,
3. $\mathsf{FSU}^\infty \mid\!\frac{\alpha}{\star}\; \Gamma_{\vec{x}} \quad \Rightarrow \quad$ For all $\beta > 0$ we have: $\mathcal{M}(\beta, \beta + 2^\alpha) \models \Gamma_{\vec{n}}$.

These bounds does not change if we strengthen FSU by (U-Lin) or (U-Nor).

Proof: In all three cases, the proof is by induction on the length of the derivation in the Tait calculus. For the verification of $(\mathsf{C\text{-}I_N})^T$ and $(\mathsf{T\text{-}I_N})^T$, we follow nearly literally the procedure of Cantini in [Can89], cf. also [Can96]. In the following we drop the indices $_{\vec{x}}$ and $_{\vec{n}}$ for the sake of readability.

1. $\mathsf{FSU}^T + (\mathsf{C\text{-}I_N})^T$

   Induction base. If $k = 0$, then one element of Γ is an axiom. Of course, the logical axioms and the axioms from TON are satisfied. From the axioms of FSU we will check the basic ones, the closure for universal quantification, the order axiom and the limit axioms.

   (1) $\Gamma, \widetilde{\mathsf{U}}(s), \mathsf{T}(s\,t) \vee \mathsf{T}(\dot{\neg}\,(s\,t))$.

   Let $l < m$ be minimal such that $\mathsf{U}_{l+1}(s)$ holds. Now by case distinction on $\mathsf{T}_l(t) \vee \neg\mathsf{T}_l(t)$ we get by the clauses 11. and 12. $\mathsf{T}_{l+1}(s\,t) \vee \mathsf{T}_{l+1}(s\,t)$. Finally, persistence yields $\mathsf{T}_{m+1}(s\,t) \vee \mathsf{T}_{m+1}(s\,t)$.

   (2) $\Gamma, \widetilde{\mathsf{U}}(s), \widetilde{\mathsf{T}}(s\,t), \mathsf{T}(t)$.

   Let $l < m$ be minimal such that $\mathsf{U}_{l+1}(s)$ and $\mathsf{T}_{l+1}(s\,t)$ holds. By clause 11. we get $\mathsf{T}_l(t)$ and persistence yields $\mathsf{T}_{m+1}(t)$.

   (10a) $\Gamma, \widetilde{\mathsf{U}}(s), (\exists x.\widetilde{\mathsf{T}}(s\,(t\,x))), \mathsf{T}(s\,(\dot{\forall}\,t))$.

   Let $l < m$ be minimal such that $\mathsf{U}_{l+1}(s)$ and $\mathsf{T}_{l+1}(s\,(t\,c))$ for all c. Clause 11. yields $\mathsf{T}_l(t\,c)$ for each c. By clause 9. we have $\mathsf{T}_l(\dot{\forall}\,t)$ and, again by 11., $\mathsf{T}_n(s(\dot{\forall}\,t))$. Persistence finally yields $\mathsf{T}_{m+1}(s\,(\dot{\forall}\,t))$.

   Note that for the adaptation of this step in the transfinite cases, it is important that we build universes on successor levels only. Otherwise, we could not find one fixed level on which $\mathsf{T}(t\,c)$ holds for all c.

(10b) $\Gamma, \widetilde{\mathsf{U}}(s), \widetilde{\mathsf{T}}(s\,(\dot{\forall}\,t)), \forall x.\mathsf{T}(s\,(t\,x))$.

Let $l < m$ be minimal such that $\mathsf{U}_{l+1}(s)$ and $\mathsf{T}_{l+1}(s\,(\dot{\forall}\,t))$. By clause 11. we have $\mathsf{T}_l(\dot{\forall}\,t)$ and by 9. $\mathsf{T}_l(t\,c)$ for all c. Thus, by clause 11. we have $\mathsf{T}_{l+1}(s\,(t\,c))$ for each c, i.e. together with persistence $\forall x.\mathsf{T}_{m+1}(s\,(t\,x))$.

(12) $\Gamma, \widetilde{\mathsf{U}}(r), \widetilde{\mathsf{U}}(s), \widetilde{\mathsf{T}}(s\,(r\,t)), \mathsf{T}(s\,t)$.

Let $l < m$ be minimal such that $\mathsf{U}_{l+1}(s)$ and $\mathsf{T}_{l+1}(s\,(r\,t))$. By clause 11. we get $\mathsf{T}_l(r\,t)$ and let $k \leq l$ minimal such that $\mathsf{T}_k(r\,t)$. On the other hand, there is $n < m$ minimal such tat $\mathsf{U}_{n+1}(r)$, i.e. $\exists b.r = \ell\,b \wedge \mathsf{C}_n(b)$. Therefore, according to clause 11. $\mathsf{T}_k(r\,t)$ can become true only when $k = n+1$. Thus, we have $\mathsf{U}_{n+1}(r)$ and $\mathsf{T}_{n+1}(r\,t)$ for $n+1 < l$ and clause 11. yields $\mathsf{T}_n(t)$. By persistence we have $\mathsf{T}_l(t)$ and, since $\mathsf{U}_{l+1}(s)$ holds, we get $\mathsf{T}_{l+1}(s\,t)$. Finally, persistence yields $\mathsf{T}_{m+1}(s\,t)$.

(13) $\Gamma, \widetilde{\mathsf{U}}(s), \widetilde{\mathsf{T}}(s\,t), \widetilde{\mathsf{T}}(s\,(\dot{\neg}\,t))$.

Let $l < m$ be minimal such that $\mathsf{U}_{l+1}(s)$ and $\mathsf{T}_{l+1}(s\,t)$. Thus, we have $\mathsf{T}_l(t)$, i.e. $t \in X_l$. Assume $\mathsf{T}_{l+1}(s\,(\dot{\neg}\,t))$. This would imply $\mathsf{T}_l(\dot{\neg}\,t)$, i.e. $\dot{\neg}\,t \in X_l$. But this contradicts the consisteny of X_l.

(14a) $\Gamma, (\exists x.\widetilde{\mathsf{T}}(t\,x) \wedge \widetilde{\mathsf{T}}(\dot{\neg}\,(t\,x))), \mathsf{U}(\ell\,t)$.

Let $l < m$ be minimal such that $\forall x.\mathsf{T}_l(t\,x) \vee \mathsf{T}_l(\dot{\neg}\,(t\,x))$, i.e. $\mathsf{C}_l(t)$. Therefore, we have $\mathsf{U}_{l+1}(\ell\,t)$ and with persistence, $\mathsf{U}_{m+1}(\ell\,t)$.

(14b) $\Gamma, (\exists x.\widetilde{\mathsf{T}}(t\,x) \wedge \widetilde{\mathsf{T}}(\dot{\neg}\,(t\,x))), \widetilde{\mathsf{T}}(t\,s), \mathsf{T}(\ell\,t\,(t\,s))$.

Let $l < m$ be minimal such that $\mathsf{C}_l(t)$ and $\mathsf{T}_l(t\,s)$. Thus we have $\mathsf{U}_{l+1}(\ell\,t)$ and clause 11 yields $\mathsf{T}_{l+1}(\ell\,t\,(t\,s))$. Again, with persistence we get $\mathsf{T}_{m+1}(\ell\,t\,(t\,s))$.

(14c) Follows analogously.

Induction step. We need to check the cut rule and class induction.

(Cut) If Γ is the conclusion of a cut rule, there is a T^+ formula φ and a natural number $l < k$ such that

$$\begin{array}{rcl} \mathsf{FSU}^T + (\mathsf{C\text{-}I_N})^T & \vdash^{l}_{\star} & \Gamma, \varphi, \\ \mathsf{FSU}^T + (\mathsf{C\text{-}I_N})^T & \vdash^{l}_{\star} & \Gamma, \neg\varphi. \end{array}$$

By the induction hypothesis used with m in the first case and $m + 2^l$ in the second, we have

$$\begin{array}{rcl} \mathcal{M}(m, m+2^l) & \models & \Gamma, \varphi, \\ \mathcal{M}(m+2^l, m+2^l+2^l) & \models & \Gamma, \neg\varphi. \end{array}$$

Since for $l < k$, $m + 2^l + 2^l \leq m + 2^k$, we have by persistence:

$$\begin{array}{rcl} \mathcal{M}(m, m+2^k) & \models & \Gamma, \varphi, \\ \mathcal{M}(m, m+2^k) & \models & \Gamma, \neg\varphi, \end{array}$$

which yields the conclusion.

$(\mathsf{C\text{-}I_N})^T$ From the premise of class induction, we know that there is a $l < k$ such that:

$$\begin{array}{rcl} \mathsf{FSU}^T + (\mathsf{C\text{-}I_N})^T & \vdash^{l}_{\star} & \Gamma, \mathsf{C}(t), \\ \mathsf{FSU}^T + (\mathsf{C\text{-}I_N})^T & \vdash^{l}_{\star} & \Gamma, \mathsf{T}(t\,0), \\ \mathsf{FSU}^T + (\mathsf{C\text{-}I_N})^T & \vdash^{l}_{\star} & \Gamma, \forall x : \mathsf{N}.\mathsf{T}(\dot{\neg}\,(t\,x)) \vee \mathsf{T}(t\,(\mathsf{s_N}\,x)). \end{array}$$

By the induction hypothesis, it follows for all $m > 0$:

$$\begin{aligned} \mathcal{M}(m, m+2^l) &\models \Gamma, \forall x.\mathsf{T}(t\,x) \vee \mathsf{T}(\dot\neg\,(t\,x)), & (1)\\ \mathcal{M}(m, m+2^l) &\models \Gamma, \mathsf{T}(t\,0), & (2)\\ \mathcal{M}(m, m+2^l) &\models \Gamma, \forall x \in \mathcal{N}.\mathsf{T}(\dot\neg\,(t\,x)) \vee \mathsf{T}(t\,(x+1)). & (3) \end{aligned}$$

By use of induction on the natural numbers we can show for all $i \in \mathcal{N}$:

$$\mathcal{M}(m, m+2^l+2^l) \models \Gamma, \mathsf{T}_{m+2^l}(t\,i). \qquad (\star)$$

For $i = 0$, this follows from (2). If $i = j+1$, we have by the (side) induction hypothesis:

$$\mathcal{M}(m, m+2^l+2^l) \models \Gamma, \mathsf{T}_{m+2^l}(t\,j).$$

From (3), we get with persistence:

$$\mathcal{M}(m, m+2^l+2^l) \models \Gamma, \mathsf{T}_{m+2^l}(\dot\neg\,(t\,j)), \mathsf{T}_{m+2^l}(t\,(j+1)).$$

From (1), we get as a special case:

$$\mathcal{M}(m, m+2^l) \models \Gamma, \mathsf{T}_{m+2^l}(t\,j), \mathsf{T}_{m+2^l}(\dot\neg\,(t\,j)).$$

By consistency of X_{m+2}, and persistence, we have therefore

$$\mathcal{M}(m, m+2^l+2^l) \models \Gamma, \neg\mathsf{T}_{m+2^l}(\dot\neg\,(t\,j)), \neg\mathsf{T}_{m+2^l}(t\,j),$$

With two cuts we get, finally,

$$\mathcal{M}(m, m+2^l+2^l) \models \Gamma, \mathsf{T}_{m+2^l}(t\,(j+1)).$$

So $(\star)$ is proven and we get

$$\mathcal{M}(m, m+2^l+2^l) \models \Gamma, \forall x \in \mathcal{N}.\mathsf{T}_{m+2^l}(t\,x),$$

which yields the required conclusion

$$\mathcal{M}(m, m+2^l+2^l) \models \Gamma, \forall x \in \mathcal{N}.\mathsf{T}(t\,x).$$

2. $\mathsf{FSU}^T + (\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})^T$

The induction base is completely analogous to the previous case. In the induction step, cut follows like in the previous proof, and we have only to check truth induction.

$(\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})^T$ From the premise of truth induction, we know that there is a $l < k$ such that:

$$\begin{aligned} \mathsf{FSU}^T + (\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})^T &\mathrel{\vdash^{l}_{\star}} \Gamma, \mathsf{T}(t\,0),\\ \mathsf{FSU}^T + (\mathsf{T}\text{-}\mathsf{I}_\mathsf{N})^T &\mathrel{\vdash^{l}_{\star}} \Gamma, \forall x : \mathsf{N}.\mathsf{T}(t\,x) \to \mathsf{T}(t\,(\mathsf{s}_\mathsf{N}\,x)). \end{aligned}$$

By the induction hypothesis, it follows for all $m > 0$:

$$\begin{aligned} \mathcal{M}(\beta, \beta+\omega^l) &\models \Gamma, \mathsf{T}(t\,0), & (1)\\ \mathcal{M}(\beta, \beta+\omega^l) &\models \Gamma, \forall x \in \mathcal{N}.\neg\mathsf{T}(t\,x) \vee \mathsf{T}(t\,(x+1)). & (2) \end{aligned}$$

By induction on i, we prove that

$$\mathcal{M}(\beta, \beta+\omega^l(i+1)) \models \Gamma, \mathsf{T}(t\,i). \qquad (\star)$$

For $i = 0$, this is (1). So let $i = j + 1$. From (2), we get

$$\mathcal{M}(\beta + \omega^l(j+1), \beta + \omega^l(j+2)) \quad \models \quad \Gamma, \neg \mathsf{T}(t\,j), \mathsf{T}(t\,(j+1)).$$

With persistence, we have

$$\mathcal{M}(\beta, \beta + \omega^l(j+2)) \quad \models \quad \Gamma, \neg \mathsf{T}_{\beta+\omega^l(j+1)}(t\,j), \mathsf{T}_{\beta+\omega^l(j+2)}(t\,(j+1)).$$

Using the (side) induction hypothesis

$$\mathcal{M}(\beta, \beta + \omega^l(j+1)) \models \Gamma, \mathsf{T}(t\,j),$$

we get with persistence

$$\mathcal{M}(\beta, \beta + \omega^l(j+2)) \models \Gamma, \mathsf{T}_{\beta+\omega^l(j+1)}(t\,j),$$

and a cut yields the required conclusion of the side induction:

$$\mathcal{M}(\beta, \beta + \omega^l(j+2)) \quad \models \quad \Gamma, \mathsf{T}(t\,(j+1)).$$

Since $\beta + \omega^l(j+1) < \beta + \omega^{l+1} \leq \beta + \omega^k$, we get from ($\star$) by persistence:

$$\mathcal{M}(\beta, \beta + \omega^k) \models \mathsf{T}(t\,i)$$

for arbitrary i in $\mathcal{N}$. So we have

$$\mathcal{M}(\beta, \beta + \omega^k) \models \forall x \in \mathcal{N}.\mathsf{T}(t\,x).$$

3. FSU^{∞}: The proof follows the standard patterns, similar to the previous ones, using the transfinite bounds.

A verification of (U-Lin) is an easy exercise following from the linearity of the ordinals along which the structures $\mathcal{M}(\gamma)$ are build. (U-Nor) is a direct consequence from the interpretation of U. □

The operator $\Phi(X, Y, x, \alpha)$ was defined in a way that it can be directly translated into $\widehat{\mathsf{ID}}_\alpha$. However, as for FON we have to introduce auxiliary theories $\Phi^+\text{-}\widehat{\mathsf{ID}}_\alpha$ which involves an additional axiom for the consistency of the fixed points of Φ. This theory is as $\widehat{\mathsf{ID}}_\alpha$ formulated in the language $\mathcal{L}_{\mathsf{fix}}$ and we use the same abbreviations. Let the operator form Φ be given as above. Now, we define $\Phi^+\text{-}\widehat{\mathsf{ID}}_\alpha$ as the extension of PA together with induction on the natural numbers for all $\mathcal{L}_{\mathsf{fix}}$ formulae by the following axioms for the new constant $\mathcal{P}^\Phi$:

$$\forall \beta \prec \alpha. \forall x. \mathcal{P}^\Phi_\beta(x) \leftrightarrow \Phi(\mathcal{P}^\Phi_\beta, \mathcal{P}^\Phi_{\prec\beta}, x, \beta),$$
$$\forall \beta \prec \alpha. \forall x. \neg(\mathcal{P}^\Phi_\beta(x) \wedge \mathcal{P}^\Phi_\beta(\dot{\neg}\, x)).$$

With the usual proof-theoretic machinery, we can formalize the assymetric interpretation in $\Phi^+\text{-}\widehat{\mathsf{ID}}_\alpha$ and get the following proposition.

Proposition 32

1. $\mathsf{FSU} + (\mathsf{C}\text{-}\mathsf{I_N})$ can be embedded in $\Phi^+\text{-}\widehat{\mathsf{ID}}_{<\omega}$.
2. $\mathsf{FSU} + (\mathsf{T}\text{-}\mathsf{I_N})$ can be embedded in $\Phi^+\text{-}\widehat{\mathsf{ID}}_{<\omega^\omega}$.
3. $\mathsf{FSU} + (\mathcal{L}_U\text{-}\mathsf{I_N})$ can be embedded in $\Phi^+\text{-}\widehat{\mathsf{ID}}_{<\varepsilon_0}$.

All three embeddings preserve (the translations of) the arithmetical theorems.

By a slight modification of the proof-theoretic analysis of the theories $\widehat{\mathsf{ID}}_\alpha$ given in [JKSS99] one can check that the theories Φ^+-$\widehat{\mathsf{ID}}_\alpha$ indeed have the same strength. The analysis of $\widehat{\mathsf{ID}}_\alpha$ uses auxiliary semi-formal systems H_α. They are essentially Tait-style calculi for $\widehat{\mathsf{ID}}_\alpha$ comprising an ω rule for the natural numbers. For our purpose we extend H_α to H_α^+ by additional axioms stating the consistency of the fixed points of the particular operator form Φ, i.e. we have to add axioms of the form

$$\neg(P^\Phi_\beta(s) \wedge P^\Phi_\beta(\dot\neg\, s))$$

for all $\beta < \alpha$. These axioms do not affect the partial cut elimination as it is done for H_α. However, in the proof of the so-called Main Lemma II, [JKSS99, lemma 13], one has to carry through "the standard elimination procedure of finitely many fixed points (say, by asymmetric interpretation, cf. e.g. [Can85, JS96, MS98])". This elimination procedure for the highest fixed points has to be adjusted in a similar way like the analysis of $\mathsf{FON} + (\mathcal{L}_F\text{-}\mathrm{I}_\mathsf{N})$ described above. That means we have to add an intermediate step where we remove the consistency axiom for the highest fixed point by use of an auxiliary system providing ordinals and Δ^Ω_0-induction on these ordinals. But this can be added without "proof-theoretic costs". So, partial cut-elimination and an asymmetric interpretation into the lower semi-formal system will still yield the appropriate bounds.

Collecting all together, in particular the last considerations, proposition 26 and theorem 24 we get the final result about the proof-theoretic strength of our theories:

Theorem 33

1. $|\mathsf{FSU} + (\mathsf{C}\text{-}\mathrm{I}_\mathsf{N})\,| = |\widehat{\mathsf{ID}}_{<\omega}| = \varphi 100 = \Gamma_0$,
2. $|\mathsf{FSU} + (\mathsf{T}\text{-}\mathrm{I}_\mathsf{N})\,| = |\widehat{\mathsf{ID}}_{<\omega^\omega}| = \varphi 1\omega 0$,
3. $|\mathsf{FSU} + (\mathcal{L}_U\text{-}\mathrm{I}_\mathsf{N})\,| = |\widehat{\mathsf{ID}}_{<\varepsilon_0}| = \varphi 1\varepsilon_0 0$.

Acknowledgment

The presented work is an improved and extended version of the third part of my dissertation [Kah97a], supervised by G. Jäger (Bern), with support of the Schweizerische Nationalfonds. I thank Andrea Cantini, Gerhard Jäger, Markus Marzetta, Anton Setzer and an anonymous referee for valuable remarks and suggestions. Special thanks go to Thomas Strahm for his invaluable help. In particular, he suggested the given treatment of the consistency axiom in the proof of the upper bounds. The revised version of this paper was prepared during a stay at the Mittag-Leffer institute, Djursholm, Sweden and I am grateful for the generous support of the institute.

References

[Acz77] Peter Aczel. The strength of Martin Löf's intuitionistic type theory with one universe. In Seppo Miettinen and Jonko Väänänen, editors, *Proceedings of the Symposiums on Mathematical Logic in Oulo 1974 and in Helsinki 1975*. Report No. 2 from the Dept. of Philosophy, University of Helsinki, 1977.

[Acz80] Peter Aczel. Frege structures and the notion of proposition, truth and set. In J. Barwise, H. Keisler, and K. Kunen, editors, *The Kleene Symposium*, pages 31–59. North-Holland, 1980.

[Bar75] Jon Barwise. *Admissible Sets and Structures*. Spinger, 1975.

[Bee85] Michael Beeson. *Foundations of Constructive Mathematics*. Ergebnisse der Mathematik und ihrer Grenzgebiete; 3.Folge, Bd. 6. Springer, 1985.

[BFPS81] Wilfried Buchholz, Solomon Feferman, Wolfram Pohlers, and Wilfried Sieg. *Iterated Inductive Definitions and Subsystems of Analysis: Recent Proof-Theoretical studies*, volume 897 of *Lecture Notes in Mathematics*. Springer, 1981.

[Can85] Andrea Cantini. A note on a predicatively reducible theory of iterated elementary induction. *Bullettino Unione Mathematica Italiana, Serie VI*, IV-B(2):413–430, 1985.

[Can86] Andrea Cantini. On the relationship between choice and comprehension principles in second order arithmetic. *Journal of Symbolic Logic*, 51:360–373, 1986.

[Can89] Andrea Cantini. Notes on formal theories of truth. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 35:97–130, 1989.

[Can90] Andrea Cantini. A theory of formal truth arithmetically equivalent to ID_1. *Journal of Symbolic Logic*, 55(1):244–259, 1990.

[Can93] Andrea Cantini. Extending the first-order theory of combinators with self-referential truth. *Journal of Symbolic Logic*, 58(2), 1993.

[Can95] Andrea Cantini. Levels of truth. *Notre Dame Journal of Formal Logic*, 36:185–213, 1995.

[Can96] Andrea Cantini. *Logical Frameworks for Truth and Abstraction*, volume 135 of *Studies in Logic and the Foundations of Mathematics*. North-Holland, 1996.

[Fef75] Solomon Feferman. A language and axioms for explicit mathematics. In J. Crossley, editor, *Algebra and Logic*, volume 450 of *Lecture Notes in Mathematics*, pages 87–139. Springer, 1975.

[Fef79] Solomon Feferman. Constructive theories of functions and classes. In M. Boffa, D. van Dalen, and K. McAloon, editors, *Logic Colloquium 78*, pages 159–224. North–Holland, 1979.

[Fef82] Solomon Feferman. Iterated inductive fixed-point theories: Application to Hancock's conjecture. In G. Metakides, editor, *Patras Logic Symposion*, pages 171–196. North–Holland, 1982.

[Fef88] Solomon Feferman. Hilbert's program relativized: Proof-theoretical and foundational reductions. *Journal of Symbolic Logic*, 53(2):364–384, June 1988.

[Fef00] Solomon Feferman. Does reductive proof theory have a viable rationale? *Erkenntnis*, 53(1–2):63–96, 2000.

[FJ93] Solomon Feferman and Gerhard Jäger. Systems of explicit mathematics with non-constructive μ-operator. Part I. *Annals of Pure and Applied Logic*, 65(3):243–263, December 1993.

[FJ96] Solomon Feferman and Gerhard Jäger. Systems of explicit mathematics with non-constructive μ-operator. Part II. *Annals of Pure and Applied Logic*, 79:37–52, 1996.

[FM87a] Robert Flagg and John Myhill. An extension of Frege structures. In D. Kueker, E. Lopez-Escobar, and C. Smith, editors, *Mathematical Logic and theoretical computer science*, pages 197–217. Dekker, 1987.

[FM87b] Robert Flagg and John Myhill. Implication and analysis in classical Frege structures. *Annals of Pure and Applied Logic*, 34:33–85, 1987.

[FS00] Solomon Feferman and Thomas Strahm. The unfolding of non-finitist arithmetic. *Annals of Pure and Applied Logic*, 104(1–3):75–96, July 2000.

[Gla93] Thomas Glaß. *Standardstrukturen für Systeme Expliziter Mathematik*. Dissertation, Westfälische Wilhelms-Universität Münster, 1993.

[Gla95] Thomas Glaß. Understanding uniformity in Feferman's explicit mathematics. *Annals of Pure and Applied Logic*, 75(1–2):89–106, 1995.

[GS96] Thomas Glaß and Thomas Strahm. Systems of explicit mathematics with non-constructive μ-operator and join. *Annals of Pure and Applied Logic*, 82(2):193–219, December 1996.

[Hin78] Peter Hinman. *Recursion-Theoretic Hierarchies*. Springer, 1978.

[HK95] Susumu Hayashi and Satoshi Kobayashi. A new formulation of Feferman's system of functions and classes and its relation to Frege structures. *International Journal of Foundations of Computer Science*, 6(3):187–202, September 1995.

[Jäg80] Gerhard Jäger. Beweistheorie von KPN. *Archiv für mathematische Logik und Grundlagenforschung*, 20:53–64, 1980.

[Jäg84] Gerhard Jäger. The strength of admissibility without foundation. *Journal of Symbolic Logic*, 49(3):867–879, September 1984.

[Jäg93] Gerhard Jäger. Fixed points in Peano arithmetic with ordinals. *Annals of Pure and Applied Logic*, 60:119–132, 1993.

[Jäg0x] Gerhard Jäger. Metapredicative and explicit Mahlo: a proof-theoretic perspective. In *Logic Colloquium 2000*, Lecture Note in Logic. A K Peters, 200x. To appear.

[JKS99] Gerhard Jäger, Reinhard Kahle, and Thomas Strahm. On applicative theories. In A. Cantini, E. Casari, and P. Minari, editors, *Logic and Foundation of Mathematics*, pages 88–92. Kluwer, 1999.

[JKS0x] Gerhard Jäger, Reinhard Kahle, and Thomas Studer. Universes in explicit mathematics. *Annals of Pure and Applied Logic*, 200x. To appear.

[JKSS99] Gerhard Jäger, Reinhard Kahle, Anton Setzer, and Thomas Strahm. The proof-theoretic analysis of transfinitely iterated fixed point theories. *Journal of Symbolic Logic*, 64(1):53–67, March 1999.

[JS95a] Gerhard Jäger and Thomas Strahm. Second order theories with ordinals and elementary comprehension. *Archive for Mathematical Logic*, 34(6):345–375, 1995.

[JS95b] Gerhard Jäger and Thomas Strahm. Totality in applicative theories. *Annals of Pure and Applied Logic*, 74:105–120, 1995.

[JS96] Gerhard Jäger and Thomas Strahm. Some theories with positive induction of ordinal strength $\varphi\omega 0$. *Journal of Symbolic Logic*, 61(3):818–842, September 1996.

[JS0xa] Gerhard Jäger and Thomas Strahm. Upper bounds for metapredicative Mahlo in explicit mathematics and admissible set theory. *Journal of Symbolic Logic*, 200x. To appear.

[JS0xb] Gerhard Jäger and Thomas Studer. Extending the system T_0 of explicit mathematics: the limit and Mahlo axioms. *Annals of Pure and Applied Logic*, 200x. To appear.

[Kah97a] Reinhard Kahle. *Applikative Theorien und Frege-Strukturen*. Dissertation, Institut für Informatik und angewandte Mathematik, Universität Bern, 1997.

[Kah97b] Reinhard Kahle. Uniform limit in explicit mathematics with universes. Technical Report IAM-97-002, IAM, Universität Bern, März 1997.

[Kah99] Reinhard Kahle. Frege structures for partial applicative theories. *Journal of Logic and Computation*, 9(5):683–700, October 1999.

[Kah0x] Reinhard Kahle. Truth in applicative theories. *Studia Logica*, 200x. To appear.

[KS0x] Reinhard Kahle and Thomas Studer. Formalizing non-termination of recursive programs. *The Journal of Logic and Algebraic Programming*, 200x. To appear.

[Mar84] Per Martin-Löf. *Intuitionistic Type Theory*. Bibliopolis, Napoli, 1984.

[Mar93] Markus Marzetta. Universes in the theory of types and names. In E. Börger et al., editor, *Computer Science Logic '92*, volume 702 of *Lecture Notes in Computer Science*, pages 340–351. Springer, 1993.

[Mar94] Markus Marzetta. *Predicative Theories of Types and Names*. Dissertation, Universtiät Bern, Institut für Informatik und angewandte Mathematik, 1994.

[MS98] Markus Marzetta and Thomas Strahm. The μ quantification operator in explicit mathematics with universes and iterated fixed point theories with ordinals. *Archive for Mathematical Logic*, 37(5–6):391–413, 1998.

[Poh89] Wolfram Pohlers. *Proof Theory*, volume 1407 of *Lecture Notes in Mathematics*. Springer, 1989.

[Rat00] Michael Rathjen. The strength of Martin-Löf type theory with a superuniverse. Part I. *Archive for Mathematical Logic*, 39(1):1–39, 2000.

[Rat0x] Michael Rathjen. The strength of Martin-Löf type theory with a superuniverse. Part II. *Archive for Mathematical Logic*, 200x. To appear.

[Rüe00] Christian Rüede. *Metapredicative Subsystems of Analysis*. Dissertation, Universität Bern, Institut für Informatik und angewandte Mathematik, 2000.

[Rüe0xa] Christian Rüede. Metapredicative subsystems of analysis. 200x. Submitted.

[Rüe0xb] Christian Rüede. Universes in metapredicative analysis. 200x. Submitted.

[Sch77] Kurt Schütte. *Proof Theory*, volume 225 of *Grundlehren der mathematischen Wissenschaften*. Springer, 1977.

[Sco75] Dana Scott. Combinators and classes. In C. Böhm, editor, *λ-Calculus and Computer Science Theory*, volume 37 of *Lecture Notes in Computer Science*, pages 1–26. Springer, 1975.

[Set99] Anton Setzer. Ordinal systems. In S. Cooper and J. Truss, editors, *Sets and Proofs*, volume 258 of *London Mathematical Society Lecture Note Series*, pages 301–338. Cambridge University Press, 1999.

[Sim99] Stephen Simpson. *Subsystems of Second Order Arithmetic*. Perspectives in Mathematical Logic. Springer, 1999.

[Str96] Thomas Strahm. *On the proof theory of applicative theories*. Dissertation, Universität Bern, Institut für Informatik und angewandte Mathematik, 1996.

[Str99] Thomas Strahm. First steps into metapredicativity in explicit mathematics. In S. Cooper and J. Truss, editors, *Sets and Proofs*, volume 258 of *London Mathematical Society Lecture Note Series*, pages 383–402. Cambridge University Press, 1999.

[Str00a] Thomas Strahm. Autonomous fixed point progressions and fixed point transfinite recursion. In Sam Buss and Pavel Pudlák Petr Hájek, editors, *Logic Colloquium '98*, volume 13 of *Lecture Notes in Logic*. A K Peters, 2000.

[Str00b] Thomas Strahm. The non-constructive μ-operator, fixed point theories with ordinals, and the bar rule. *Annals of Pure and Applied Logic*, 104(1–3):305–324, July 2000.

[Tai68] William Tait. Normal derivability in classical logic. In J. Barwise, editor, *The Syntax and Semantic of Infinitary Languages*, volume 72 of *Lecture Notes in Mathematics*, pages 204–236. Springer, 1968.