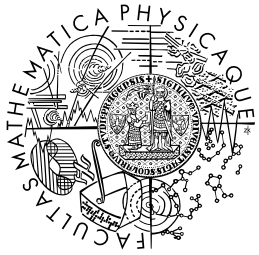

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

DOCTORAL THESIS

Amirhossein Akbar Tabatabai

In the Light of Intuitionism: Two Investigations in Proof Theory

Department of Algebra

Supervisor of the doctoral thesis: Prof. RNDr. Pavel Pudlák, DrSc

Study programme: Mathematics

Study branch: Algebra, Theory of Numbers and Mathematical Logic

Prague 2018

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ..... signature of the author

Acknowledgment. This work has been supported by the institute of mathematics of the Czech academy of sciences and the ERC advanced project FEALORA, lead by prof. Pavel Pudlák.

First and foremost, I like to thank Pavel Pudlák to whom I am incredibly indebted, for his guidance, his support, the discussions that we have had and all the things I have learned from him. I also wish to thank Jan Krajiček from whom I learned a lot and Lev Beklemishev, Emil Jeřábek, Pavel Hrubeš and Neil Thapen for reading some parts of the earlier draft and pointing out some errors in the earlier proofs and arguments. Last but not least, I would like to thank Mohammad Ardeshir, Arnold Beckmann, Sam Buss, Anna Horská, Raheleh Jalali and Masoud Memarzadeh for their support and their helpful discussions and comments.

Title: In the Light of Intuitionism: Two Investigations in Proof Theory

Author: Amirhossein Akbar Tabatabai

Department: Department of Algebra

Supervisor: Prof. RNDr. Pavel Pudlák, DrSc, Institute of Mathematics, Academy of Sciences of the Czech Republic

Abstract: This dissertation focuses on two specific interconnections between the classical and the intuitionistic proof theory. In the first part, we will propose a formalization for Gödel's informal reading of the BHK interpretation, using the usual classical arithmetical proofs. His provability interpretation of the propositional intuitionistic logic, first appeared in [12], in which he introduced the modal system, **S4**, as a formalization of the intuitive concept of provability and then translated **IPC** to **S4** in a sound and complete manner. His work suggested the search for a concrete provability interpretation for the modal logic **S4** which itself leads to a concrete provability interpretation for the intutionistic logic. In the first chapter of this work, we will try to solve this problem. For this purpose, we will generalize Solovay's provability interpretation of the modal logic **GL** to capture other modal logics such as **K4**, **KD4** and **S4**. Then, using the mentioned Gödel's translation, we will propose a formalization for the BHK interpretation via classical proofs. As a consequence, it will be shown that the BHK interpretation is powerful enough to admit many different formalizations that surprisingly capture different propositional logics, including intuitionistic logic, minimal logic and Visser-Ruitenburg's basic logic. We will also present some negative results to show that there is no provability interpretation for any extension of the system **KD45** and as we expected there is no BHK interpretation for the classical propositional logic.

In the second half of the dissertation, we change our focus to the other direction of the interconnection to investigate the applications of the intuitionistic viewpoint in the realm of classical proof theory. For this purpose, we will develop a complexity sensitive version of the classical Dialectica interpretation to deal with the bounded theories of arithmetic. More precisely, we will define a notion called the computational flow which is a pair consisting of a sequence of computational problems of a certain sort and a sequence of computational reductions among them. We will develop a theory for these flows to provide a sound and complete interpretation for bounded theories of arithmetic. This property helps us to transform a first order arithmetical proof to a sequence of computational reductions by which we can extract the computational content of low complexity statements in some bounded theories of arithmetic including $I\Delta_0$, T^k_n, $I\Delta_0(\exp)$ and PRA. Then, in the last section, by generalizing term-length flows to ordinal-length flows, we will extend our investigations from bounded theories to strong unbounded systems such as PA and $\mathrm{PA} + \mathrm{TI}(\alpha)$ to capture their total NP search problems.

Keywords: Provability Interpretation, BHK Interpretation, Proof Mining, Bounded Arithmetic

Contents

1. Provability Interpretation of Propositional and Modal logics

1.1 Introduction

1.1.1 BHK Interpretation

In the intuitionistic tradition, mathematics is considered as a theory of mental constructions and hence, truth naturally means the existence of a proof. Thus, provability is the cornerstone of the whole intuitionistic paradigm. With this fact in mind, like any other logic, the intuitionistic logic would be a calculus to describe the behavior of truth, which in this case, is the concept of provability. In other words, intuitionistic logic is a meta-theory of the concept of provability. Let us explain the role of connectives in this logic. Again, like any other logic, a connective is an operation on the truth content of its inputs, which in the case of intuitionistic logic means the operations on the proofs. If we want an intuitive semantics for intuitionistic logic, we have to find out what the meaning of a connective is. The answer to this question is the well-known BHK interpretation. Its propositional part is the following:

• a proof for $A \wedge B$ is a pair of a proof for A and a proof for B.
• a proof for $A \vee B$ is a proof for A or a proof for B.
• a proof for $A \to B$ is a construction which transforms any proof of A to a proof for B.
• a proof for $\neg A$ is a construction which transforms any proof of A to a proof for $\bot$.
• $\bot$ does not have any proof.

Clearly, what we presented as the BHK interpretation is just an informal interpretation and we need to find its exact formalization if we want to use it as a mathematical tool. For instance, if we want to establish an argument which shows that Heyting's formalization of **IPC** is an adequate formalization of intuitionistic viewpoint, we have to prove the soundness and completeness of **IPC** with respect to the BHK interpretation and this obviously needs an exact formalizion. Now, to formalize the interpretation, we firstly need a formalization of the concept of proof. Based on the extensive works in proof theory that have been done so far, it seems quite possible to find an appropriate formalization of the concept of proof and hence of the BHK interpretation. But, unfortunately, despite all the attempts that have been made, the BHK interpretation has not been formalized so far (for an extensive history of the problem see [2]). Why does this natural and simple interpretation resist formalization? To find an answer to this question, let us review one of the key properties of the interpretation. Think of a proposition $A \to B$. Its proof is a construction that transforms any proof of A to a proof of B. It is clear that this construction would be a meta-proof and not just a proof, because it talks about proofs and therefore it should belong to the meta-language of A and B. In other words, we could claim that the act of

introducing an implication increases the layer of the meta-language which we are arguing in. Therefore, in BHK interpretation all levels of our meta-languages are involved and this is the reason why this interpretation is so complex to formalize. Since we need to formalize the meaning of proof, we have to extend our task to find a meaning of a proof at any level of the meta-languages.

There are two different approaches to implement this idea. In the first approach, we could be faithful to the intuitionistic paradigm and find an intuitionistically valid interpretation of the proofs. However, in the second approach we could change our viewpoint and construct a bridge to find an appropriate classical interpretation of the concept of a proof to formalize the BHK interpretation. The first approach is Heyting's approach and the second one is Kolmogorov's. At first glance, the first approach seems very natural to try but there is a huge problem there; a conceptual vicious circle which forces us to understand the semantics of the paradigm, the BHK interpretation, in terms of itself and it makes the whole process very complicated. We want to emphasize that this vicious circle does not mean that the first approach is philosophically invalid, but it just shows how complex it could be. (Think of classical logic and its semantics which is based on the classical meta-theory. This is an obvious vicious circle, but these kinds of vicious circles are the inherent properties of any paradigm in the philosophy of mathematics and we have to deal with them.) In this chapter we follow the second approach and interpret all proofs as the classical proofs in different layers of meta-languages. But this is not an easy task to do and in the forthcoming part of the Introduction we will investigate the problems in this approach.

The last thing we want to mention here is that what we are going to formalize, is actually an implicit version of the BHK interpretation, instead of the original one. In the original interpretation we interpret all the connectives as operations on explicitly specified proofs. But we could somehow eliminate the *proofs* from the interpretation and just talk about the *provability* of a sentence. For instance, the disjunction case in the original BHK interpretation transforms to the following one: $A \vee B$ is provable if A is provable or B is provable. The problem here, is the case of implication which is not reducible to a simpler one. In order to solve this problem, we need a primitive connective to formalize the concept of provability. A role which would be played by the connective "box" in modal logics and this is one of the most important contributions to the problem, which was made by Kurt Gödel. Now, Gödel's contribution.

Gödel's Translation

In 1933 [12], Gödel introduced a provability interpretation of **IPC** that can be seen as an implicit version of the well-known BHK interpretation of the intuitionistic logic. By this interpretation he could justify the fact that Heyting's formalization of **IPC** is sound and complete for its intended semantics which is the BHK interpretation. Let us review some steps of his work.

1. Giving a proof interpretation: Before giving any provability interpretation of **IPC**, we should explain our intention of the concept of *provability* and the properties that we want to have. As you expect, Gödel began his work exactly

from this point. He used the language of modal logics, in which the symbol "$\Box$" is interpreted as a *provability* predicate. In the next step, he formalized the expected properties of this provability predicate by some axioms which make up the well-known modal system **S4**. Notice that in contrast with using a concrete interpretation of provability, he used a theory for formalizing this concept (**S4**). In fact, his system just characterizes the properties of our intuitive provability predicate by some formal system, and is totally silent about its real nature.
After this introduction, we are ready to give the definition of his interpretation. Consider the translation function $b : \mathcal{L} \to \mathcal{L}_{\Box}$ as follows:
$\mathcal{L}$ and $\mathcal{L}_{\Box}$ are the languages of **IPC** and **S4** respectively. [1]

(i) $p^b = \Box p$ and $\bot^b = \Box\bot$

(ii) $(A \wedge B)^b = A^b \wedge B^b$

(iii) $(A \vee B)^b = A^b \vee B^b$

(iv) $(A \to B)^b = \Box(A^b \to B^b)$

(v) $(\neg A)^b = \Box(A^b \to \Box\bot)$

It is clear that A^b is the implicit BHK interpretation of A. In fact, the definition of b is the natural paraphrase of the original BHK interpretation in terms of provability instead of proofs.
It is time to investigate the soundness-completeness property of the interpretation.

2. Soundness and Completeness: Consider the following theorem:

Theorem 1.1.1. *For any proposition $A \in \mathcal{L}$,* **IPC** $\vdash A$ *iff* **S4** $\vdash A^b$.

Proof. For the complete investigation of this theorem and some related results[2] see [12]. □

We have the system **S4** which formalizes what we expect from a provability predicate and based on the mentioned soundness-completeness result we can reduce the problem of finding a formalization of the implicit BHK interpretation to the problem of finding a provability interpretation for **S4**. Therefore, our task will be to find a concrete interpretation of this provability predicate (the connective box) in terms of classical provability in classical theories. But, consider the fact that the problem of finding a provability interpretation for **S4** has its own importance itself, independent of its relation to the BHK interpretation.

[1] In fact, our translation is different from the translation of the paper [12]. The differences are the following: $p^b = p$, $\bot^b = \bot$, $(A \to B)^b = \Box A^b \to \Box B^b$, and $(\neg A)^b = \neg\Box A^b$. While both of these two translations basically do the same task, we use the first one, because it is more compatible with our intuition of intuitionistic semantics and it is adequate for the systems weaker than **S4**.

[2] While this theorem is the heart of Gödel's work, he only stated it and left it without any proof. The soundness part is an easy consequence of induction on the length of the proof, but the completeness part was finally proved in 1947 by Tarski and McKinsey using the algebraic semantics for **S4**.

The first attempt to find a concrete provability interpretation for **S4** was made by Gödel himself. In a very negative way, he showed that the natural expected interpretation of the provability predicate is not sound for **S4**. Let us explain his result in more detail:
The most natural choice to interpret the box operator is the provability predicate of a formal theory [3]. Let T be a formal system; therefore, the meaning of $\Box A$ would be $\mathrm{Pr}_T(A)$ such that $\mathrm{Pr}_T(\cdot)$ is a provability predicate for T. (Notice that in this case we suppose our formal system T to be sufficiently strong to be able to formalize some parts of the meta-mathematics.) Consider the theorem $\Box\neg\Box\bot$ of **S4**. Its interpretation is $\mathrm{Pr}_T(\neg\mathrm{Pr}_T(\bot))$ and if it were true we would have $T \vdash \neg\mathrm{Pr}_T(\bot)$ which contradicts Gödel's second incompleteness theorem.
Therefore, we know that on the one hand, the seemingly natural way to formalize the concept of proof and provability in the BHK interpretation is to fix a formal system and interpret all the proofs as the proofs in that theory. And on the other hand, the logic **S4** is not sound with respect to this natural interpretation. This is for the case of **S4**. However, we could claim that the natural formalization of the BHK interpretation is not sound either. For instance, if you try to interpret the sentence $A \wedge (A \to B) \to B$ of intuitionistic logic, you find out that it is more or less the same as the modal formula $\Box(\Box p \to p)$ and hence intuitionistic logic inherits the same problem. In sum, we can say that the natural formalization of the BHK interpretation and also the natural interpretation of **S4** do not work. Based on these observations, we have intuition why finding a formalization of the BHK interpretation is a difficult task.

There is a natural question to ask. If the theory **S4** is intuitively valid and we know that we can not interpret the box as a provability predicate in some formal system, then what could be a natural provability interpretation of **S4**? Unfortunately, despite a lot of attempts which have been made so far, this question remains open. For instance, Kripke [15] introduced a provability interpretation which is based on his Kripke models and just captures our provability intuition for formulas without nested modalities. Or in [9], Buss introduced the "pure provability" which have the same problem with the nested modalities. Actually, the only successful attempt to find a provability interpretation is Artemov's "logic of proofs" which is based on the idea of introducing all explicit proofs, investigating the intended behavior of proofs in a theory (logic of proofs) and then interpreting the box as the existence of the proof. These explicitly mentioned proofs could empower us to avoid non-standard proofs which have the main role in Gödel's second incompleteness theorem and some of the counter-intuitive theorems in meta-mathematics. In Section 9 we will come back to Artemov's logic of proofs and we will discuss its advantages and disadvantages.

As this long introduction shows, our main problem is to find a provability interpretation for the modal logic **S4** to formalize the BHK interpretation. In this chapter, we will try to solve this problem and in the forthcoming part of the Introduction we will sketch the idea of our semantics and our key results.

[3] The system T is formal iff the set of its consequences is recursively enumerable.

1.1.2 The Main Idea and the Main Results

Why the mentioned natural proof interpretation is not a solution to our problem? One of the possible answers is the fact that this interpretation does not distinguish between languages and meta-languages. Let us illuminate this fact by an example. Suppose p is an atom. What should be an intended interpretation of p? p is an atomic sentence about the real world, it is just a description of the world and this description is in the first level. But how about $\Box p$? The intended interpretation of this formula is the provability of p in some theory. But, what is important here, is the level of the theory and the level of this sentence. Since p is a fact about the real world, the theory in which p is proved, should be a first level theory, i.e. a theory about the world. However, the sentence $(\Box p)$ is not about the real world; it is about provability and hence it should be characterized as a sentence in the second level. Therefore, the intended meaning of this second level sentence is $\mathrm{Pr}_{T_0}(p)$. Let us ask about the interpretation of $\Box\Box p$. This is about the provability of provability of p. The first box refers to a first level theory T_0. But the second box is about the provability of the provability, which has higher order, and it means the provability should be investigated in a second level theory, T_1. The important thing is the fact that there is no reason to assume that $T_1 = T_0$. Actually, our experience in mathematical logic shows that it is genuinely important to distinguish the meta-theory and the object theory, and in some crucial cases the power of the meta-theory should be more than the theory itself. For instance, Gödel's incompleteness theorems show that to answer a very basic meta-mathematical question about the system, i.e. its consistency, we need a more powerful meta-theory. Based on these sentiments, the natural way to interpret boxes in a modal sentence is interpreting them in different theories depending on the nesting level of the individual occurrence of a box. To formalize this idea, we need two different ingredients. First, a model for the real world to interpret atoms as facts about the world and second a hierarchy of theories which plays the role of the hierarchy of the meta-theories. Hence, the intended model would be $(M, \{T_n\}_{n=0}^{\infty})$ in which M is a classical model and T_n is the theory in the n-th level of the hierarchy. (We call these models, the provability models.) Moreover, we need a way of witnessing all boxes as the provability predicates of these theories in a coherent way. This is the complex part of the formalization and we will talk about it in the next section. But for now, just think of the interpretation intuitively in the sense that any outer box should be interpreted as the provability predicate of a stronger theory. Therefore, our main result for modal logics is the following:

Theorem 1.1.2. *(i) The logic* **K4** *is sound and complete with respect to the provability interpretation in all provability models.*

(ii) The logic **KD4** *is sound and complete with respect to the provability interpretation in consistent provability models, i.e.* $(M, \{T_n\}_{n=0}^{\infty})$ *where for any* n*,* M *thinks that* T_n *is consistent and* $T_{n+1} \vdash \mathrm{Cons}(T_n)$*.*

(iii) The logic **S4** *is sound and complete with respect to the provability interpretation in all reflexive provability models, i.e.* $(M, \{T_n\}_{n=0}^{\infty})$ *where for any* n*,* M *thinks that* T_n *is sound and* $T_{n+1} \vdash \mathrm{Rfn}(T_n)$*.*

(*iv*) *The logic* **GL** *is sound and complete with respect to the provability interpretation in all constant provability models, i.e.* $(M, \{T_n\}_{n=0}^{\infty})$ *where for any* n, M *thinks that* $T_n = T_0$.

(*v*) *The logic* **GLS** *is sound and complete with respect to the provability interpretation in all sound constant provability models, i.e.* $(M, \{T_n\}_{n=0}^{\infty})$ *where for any* n, M *thinks that* T_n *is sound and* $T_n = T_0$.

(*vi*) *No extensions of the logic* **KD45** *are sound in any provability model.*

Here are some remarks about this main theorem. First of all, it shows that the use of a hierarchy of meta-theories instead of just one theory to witness the box operators could define a brand new framework to capture different modal logics depending on provability interpretation. In fact, it shows that modal logics could be seen as the formal theories to describe the relation between the real world and the theories in the hierarchy of meta-theories which we use; in other words, they are theories for the whole discourse of provability. Moreover, in the case of the logics **K4**, **KD4** and **S4** it shows that they describe the relation of the model and meta-theories in a natural and expected way. For instance, in an informal reading of the axiom $\Box A \to A$ in **S4**, we mean that our proofs are sound. And this is exactly one of the conditions we put on the models to capture the logic **S4**. It is similar for all other axioms, logics and conditions in the aforementioned result.
Secondly, the result shows that if we restrict the whole hierarchy of meta-theories to just one theory, we could reconstruct Solovay's results for **GL** and **GLS**. Therefore, it shows that our provability interpretation is a generalization of Solovay's interpretation and our main result is a generalization of Solovay's results.

If we combine this provability interpretations with Gödel translation, we will have different BHK interpretations with respect to different powers of meta-theories. We have:

Theorem 1.1.3. (*i*) *The logic* **BPC** *is sound and complete with respect to the BHK interpretation in all provability models.*

(*ii*) *The logic* **EBPC** *is sound and complete with respect to the BHK interpretation in all consistent provability models, i.e.* $(M, \{T_n\}_{n=0}^{\infty})$ *where for any* n, M *thinks that* T_n *is consistent and* $T_{n+1} \vdash \mathrm{Cons}(T_n)$.

(*iii*) *The logic* **MPC** *is sound and complete with respect to the weak BHK interpretation in all reflexive provability models, i.e.* $(M, \{T_n\}_{n=0}^{\infty})$ *where for any* n, M *thinks that* T_n *is sound and* $T_{n+1} \vdash \mathrm{Rfn}(T_n)$.

(*iv*) *The logic* **IPC** *is sound and complete with respect to the BHK interpretation in all reflexive provability models, i.e.* $(M, \{T_n\}_{n=0}^{\infty})$ *where for any* n, M *thinks that* T_n *is sound and* $T_{n+1} \vdash \mathrm{Rfn}(T_n)$.

(*v*) *The logic* **FPL** *is sound and complete with respect to the BHK interpretation in all constant provability models, i.e.* $(M, \{T_n\}_{n=0}^{\infty})$ *where for any* n, M *thinks that* $T_n = T_m$.

(*vi*) *The logic* **CPC** *does not admit any BHK interpretations.*

If you are not familiar with these propositional logics, we will define them in the Preliminaries section. But for now, just assume that the propositional logics **BPC**, **EBPC**, **IPC** and **FPL** are the propositional counterparts of the modal systems **K4**, **KD4**, **S4** and **GL**, respectively. Moreover, by weak BHK interpretation, we informally mean the usual BHK interpretation without the consistency condition. This is the last condition in the BHK interpretation which assumes that there is no proof for $\bot$. And finally, **MPC**, roughly is **IPC** without the Ex Falso rule. The rule which makes possible to prove anything from the contradiction.

Some remarks about this result are in order. First of all, it shows that there are different BHK interpretations instead of just one. This observation somehow contradicts the folklore belief and it is surprising. The reason is that the BHK interpretation just defines the meaning of a connective in terms of the provability in different levels of meta-languages. But, it is silent about what kinds of commitments we impose on our meta-theories.

Therefore, we can impose different philosophically motivated conditions on the behavior of meta-theories to capture different propositional logics, all of them valid under the BHK interpretation. For instance, we can choose the minimal possible commitment which means that there is no non-trivial condition on the hierarchy of meta-theories. Then the BHK interpretation leads to the logic **BPC**. On the other hand, if we suppose that our meta-theories are strong enough to prove the reflection principle for lower theories and all the theories are sound, then the BHK interpretation leads to the logic **IPC**. This observation shows a key fact: There is a web of different intuitionistic logics according to the BHK interpretation; the logics **IPC** and **BPC** are just two examples of these intuitionistic logics and both of them are philosophically valid. In sum, we have to talk about *intuitionistic logics* instead of *the* intuitionistic logic.
Secondly, the result shows that our framework of provability interpretations can capture different propositional logics and just like the case of modal logics, we are able to say that propositional logics are logics to describe the behavior of the real world and the hierarchy of meta-theories. This formalizes the intuitionist claim that intuitionistic mathematics is a way to talk about proofs and proofs only.
Thirdly, it is possible to define different kinds of Gödel's translation. Hence, it is possible to capture different propositional logics via these different translations. But it is important to consider that the translation we used in the above result is the valid translation to formalize the BHK interpretation and those different kinds of translations may not be rooted in the usual BHK interpretation. However, they are still provability interpretations and could be useful.

1.2 Preliminaries

In this section we will introduce some of the preliminaries that we need in the following sections. First of all, we will introduce the sequent calculi for the modal logics **K4**, **KD4** and **S4**. Then we will introduce some propositional logics such as **BPC**, **MPC** and **IPC** as the propositional counterparts of some of the modal logics and finally we will state the Solovay's completeness results.

1.2.1 Sequent Calculi for Modal Logics

Consider the following set of rules:

Axioms:

$$A \Rightarrow A \qquad \bot \Rightarrow$$

Structural Rules:

$$(wL)\ \frac{\Gamma \Rightarrow \Delta}{\Gamma, A \Rightarrow \Delta} \qquad (wR)\ \frac{\Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, A}$$

$$(cL)\ \frac{\Gamma, A, A \Rightarrow \Delta}{\Gamma, A \Rightarrow \Delta} \qquad (cR)\ \frac{\Gamma \Rightarrow \Delta, A, A}{\Gamma \Rightarrow \Delta, A}$$

$$(cut)\ \frac{\Gamma_0 \Rightarrow \Delta_0, A \qquad \Gamma_1, A \Rightarrow \Delta_1}{\Gamma_0, \Gamma_1 \Rightarrow \Delta_0, \Delta_1}$$

Propositional Rules:

$$\vee L\ \frac{\Gamma_0, A \Rightarrow \Delta_0 \qquad \Gamma_1, B \Rightarrow \Delta_1}{\Gamma_0, \Gamma_1, A \vee B \Rightarrow \Delta_0, \Delta_1} \qquad \vee R\ \frac{\Gamma \Rightarrow \Delta, A_i}{\Gamma \Rightarrow \Delta, A_0 \vee A_1}\ (i = 0, 1)$$

$$\wedge L\ \frac{\Gamma, A_i \Rightarrow \Delta}{\Gamma, A_0 \wedge A_1 \Rightarrow \Delta}\ (i = 0, 1) \qquad \wedge R\ \frac{\Gamma_0 \Rightarrow \Delta_0, A \qquad \Gamma_1 \Rightarrow \Delta_1, B}{\Gamma_0, \Gamma_1 \Rightarrow \Delta_0, \Delta_1, A \wedge B}$$

$$\rightarrow L\ \frac{\Gamma_0 \Rightarrow A, \Delta_0 \qquad \Gamma_1, B \Rightarrow \Delta_1}{\Gamma_0, \Gamma_1, A \rightarrow B \Rightarrow \Delta_0, \Delta_1} \qquad \rightarrow R\ \frac{\Gamma, A \Rightarrow B, \Delta}{\Gamma \Rightarrow \Delta, A \rightarrow B}$$

$$\neg L\ \frac{\Gamma \Rightarrow \Delta, A}{\Gamma, \neg A \Rightarrow \Delta} \qquad \neg R\ \frac{\Gamma, A \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg A}$$

Modal Rules:

$$\Box_4 R\ \frac{\Gamma, \Box\Gamma \Rightarrow A}{\Box\Gamma \Rightarrow \Box A} \qquad \Box_D R\ \frac{\Gamma, \Box\Gamma \Rightarrow}{\Box\Gamma \Rightarrow}$$

$$\Box_S R\ \frac{\Box\Gamma \Rightarrow A}{\Box\Gamma \Rightarrow \Box A} \qquad \Box L\ \frac{\Gamma, A \Rightarrow \Delta}{\Gamma, \Box A \Rightarrow \Delta}$$

The system $G(\mathbf{K4})$ is the system that consists of the axioms, structural rules, propositional rules and the modal rule $\Box_4 R$. $G(\mathbf{KD4})$ is $G(\mathbf{K4})$ plus the rule $\Box_D R$ and finally, $G(\mathbf{S4})$ is the system $G(\mathbf{K4})$ when we replace the rule $\Box_4 R$ by $\Box_S R$ and add the rule $\Box L$. All of these systems have the cut elimination property. (See [16]).

1.2.2 Propositional Logics

The next ingredient is the propositional counterparts of the usual modal logics. The intuitionistic logic **IPC** and the minimal logic **MPC** are the well-known logics in this area, but there are also some weaker systems which are very interesting in terms of the provability interpretation. For instance, we can mention the basic propositional logic **BPC** and the formal propositional logic **FPL** defined by A. Visser in [22] or the extended basic propositional logic **EBPC** defined by M. Ardeshir and B. Hesaam in [1]. To define these logics, consider the following set of rules:

Propositional Rules:

$$\wedge I\ \frac{A \qquad B}{A \wedge B} \qquad \wedge E\ \frac{A \wedge B}{A}\ \wedge E\ \frac{A \wedge B}{B}$$

$$\vee I\ \frac{A}{A \vee B}\ \vee I\ \frac{B}{A \vee B} \qquad \vee E\ \frac{A \vee B \qquad \begin{matrix}[A]\\ \mathcal{D}\\ C\end{matrix} \qquad \begin{matrix}[B]\\ \mathcal{D}'\\ C\end{matrix}}{C}$$

$$\rightarrow I\ \frac{\begin{matrix}[A]\\ \mathcal{D}\\ B\end{matrix}}{A \rightarrow B} \qquad \perp\ \frac{\perp}{A}$$

Formalized Rules:

$$(\wedge I)_f\ \frac{A \rightarrow B \qquad A \rightarrow C}{A \rightarrow B \wedge C} \qquad (\vee E)_f\ \frac{A \rightarrow C \qquad B \rightarrow C}{A \vee B \rightarrow C}$$

$$tr_f\ \frac{A \rightarrow B \qquad B \rightarrow C}{A \rightarrow C}$$

Moreover, consider the following set of rules:

$$C\ \frac{A \qquad \neg A}{\perp} \qquad R\ \frac{A \qquad A \rightarrow B}{B}$$

$$D\ \frac{}{A \vee \neg A} \qquad L\ \frac{(A \wedge (A \rightarrow B)) \rightarrow B}{A \rightarrow B}$$

The logic **BPC** is defined as the system which consists of the propositional rules and the formalized rules. Then logic **EBPC** defined as **BPC** $+ C$, logic **FPL** is defined as **BPC** $+ L$, **IPC** is defined as **BPC** $+ R$, **MPC** is defined as **IPC** without $\perp$ rule, and finally **CPC** is defined as **IPC** $+ D$.

Remark 1.2.1. *Consider the following rules:*

$$C'\ \frac{\top \rightarrow \perp}{\perp} \qquad R'\ \frac{\top \rightarrow A}{A} \qquad L'\ \frac{(\top \rightarrow A) \rightarrow A}{\top \rightarrow A}$$

It is possible to define **EBPC** *as* **BPC** $+ D'$*;* **IPC** *as* **BPC** $+ R'$ *and* **FPL** *as* **BPC** $+ L'$*. It is obvious that D', R' and L' are special cases of D, R and L, respectively. Therefore it remains to show that D', R' and L' can simulate D, R and L, respectively. The following proofs show that it is the case:*

$$\dfrac{\dfrac{\dfrac{A}{\top \rightarrow A} \qquad A \rightarrow \perp}{C'\ \dfrac{\top \rightarrow \perp}{\perp}}}{} \qquad \dfrac{\dfrac{\dfrac{A}{\top \rightarrow A} \qquad A \rightarrow B}{R'\ \dfrac{\top \rightarrow B}{B}}}{}$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{A \to \top}}{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[\top \to (A \to B)]^2}{\top \to (A \wedge (A \to B))} \quad \dfrac{\dfrac{[A]^1}{\top \to A}}{}}{\top \to B} \quad [(A \wedge (A \to B)) \to B]^3}{A \to B}}{(\top \to (A \to B)) \to (A \to B)}}{\top \to (A \to B)}}{A \to B}}{A \to ((A \to B))}}{A \to (A \wedge (A \to B))} \quad [(A \wedge (A \to B)) \to B]^3}{(A \to B)}$$

Notice that the double lines mean simple sub-proofs that we omit and $(*)$ *is the sub-proof which proves*

$$A, (\top \to (A \to B)), ((A \wedge (A \to B)) \to B) \vdash A \to B$$

1.2.3 Solovay's Theorems

In this subsection we will mention the Solovay's seminal arithmetical completeness theorems. (See [20] and [7].) They will be needed to prove some of our completeness theorems in the next sections. Note that in the case of **GL** we will state the uniform version of the completeness theorem which will have a crucial role in our proofs.

Definition 1.2.2. *Assume that* $I\Sigma_1 \subseteq T$ *is a* Σ_1*-sound arithmetical theory. By an arithmetical substitution* σ *we mean a function from the atomic formulas in the modal language to the set of arithmetical sentences. And if* $A \in \mathcal{L}_\Box$ *is a modal formula, by* A^σ *we mean an arithmetical sentence resulted by substituting atoms by* σ*, and interpreting boxes as the provability predicate of* T*.*

Theorem 1.2.3. *(i) (First Theorem) If* $\mathbf{GL} \vdash A$ *then for all arithmetical substitutions* σ*,* $I\Sigma_1 \vdash A^\sigma$*. Moreover, there is an arithmetical substitution* $*$ *such that for all modal formulas* A*, if* $T \vdash A^*$*, then* $\mathbf{GL} \vdash A$*.*

(ii) (Second Theorem) $\mathbf{GLS} \vdash A$ *iff for all arithmetical substitutions* σ*,* $\mathbb{N} \vDash A^\sigma$*.*

1.3 Provability models

In this section we will introduce a provability model as a formalization of the intuitive combination of a model and a hierarchy of theories. Then, we will define the satisfaction relation between modal formulas and provability models. And as a conclusion, we will justify our notion of provability interpretation.

1.3.1 Definitions and Examples

Suppose that we have a modal formula A, and we want to interpret any box in the formula as a provability predicate. Note that when you have two boxes in A such that one box is in the scope of the other box, our intuition forces us to accept that

the outer box talks about the provability in the meta-theory while the inner box is just capturing the provability in the lower theories. Therefore, we can claim that the natural model for the provability interpretation of modal logics is a pair of one first order structure to interpret the atoms of the language, and a hierarchy of theories to play the role of a hierarchy of meta-theories. Moreover, we choose our structure and our theories as a model and theories for arithmetic, respectively, because in these theories we have a natural way of coding the language, the meta-language, the meta-meta-language and so on. Furthermore, we suppose that all of our theories include $I\Sigma_1$ to have enough power to formalize the basic meta-mathematics of the theories. And, for the same reason we assume $M \vDash I\Sigma_1$, because we want to have the true meta-mathematical properties obviously.

Definition 1.3.1. *A provability model is a pair $(M, \{T_n\}_{n=0}^{\infty})$ where M is a model of $I\Sigma_1$ and $\{T_n\}_{n=0}^{\infty}$ is a hierarchy of arithmetical r.e. theories such that for any n, $I\Sigma_1 \subseteq T_n \subseteq T_{n+1}$ provably in $I\Sigma_1$.*

We define expansions of a modal formula.

Definition 1.3.2. *$E(A)$, the set of all expansions of A, is inductively defined as follows:*

- *If A is an atom, $E(A) = \{A\}$.*
- *If $A = B \circ C$, then $E(A) = \{D \circ E \mid D \in E(B)$ and $E \in E(C)\}$ for $\circ \in \{\wedge, \vee, \rightarrow\}$.*
- *If $A = \neg B$, then $E(A) = \{\neg D \mid D \in E(B)\}$.*
- *If $A = \Box B$, then $E(A) = \{\Box \bigvee_{i=1}^{k} D_i \mid \forall 1 \leq i \leq k,\ D_i \in E(B)\}$.*

Moreover, if Γ is a sequence of modal formulas, by a sequence of expansions of Γ, we mean a sequence such that for any formula in Γ, it has at least one of its expansions and at most finitely many of them. We will denote these sets by $\bar{\Gamma}$.

Informally speaking, an expansion of a formula A is a formula resulted by replacing any formula after a box with disjunctions of the expansions of the formula.

Example 1.3.3. *For instance, the formula $\Box(\neg\Box(\Box p \vee \Box p) \vee \neg\Box\Box(p \vee p))$ is an expansion of the formula $\Box\neg\Box\Box p$*

So far, we have justified the Definition 1.3.1. Let us investigate the intuitive meaning of the witnesses, as well. We claim that a natural interpretation is based on the interpretion of the outer boxes as meta-theories of the inner boxes. For simplicity, we call this kind of interpretation ordered interpretation. Therefore, to have an ordered interpretation we need to interpret all of the boxes in A as the provability predicates of the theories in an ordered way. And, since for any theory we have a number which shows its layer in the hierarchy, it is enough to assign a natural number to a box. Consider that if we assign n to a box, the intended meaning is that the interpretation of that box is the provability predicate for the theory T_n. This role is played by the concept of witness. In fact, a witness is just an assignment for the boxes in an ordered way.

Notation 1.3.4. *If w_is are sequences of natural numbers, by $(w_1, w_2, \ldots, w_n)$ we mean the concatenation of w_is.*

Definition 1.3.5. *Let w be a sequence of natural numbers and A be a modal formula. Then the relation $w \Vdash A$, which means w is a witness for A, is inductively defined as follows:*

- *If A is an atom, $() \Vdash A$.*
- *If $A = B \circ C$, then $(w_1, w_2) \Vdash A$ if $w_1 \Vdash B$ and $w_2 \Vdash C$ for $\circ \in \{\wedge, \vee, \rightarrow\}$*
- *If $A = \neg B$, then $w \Vdash A$ if $w \Vdash B$.*
- *If $A = \Box B$, then $(n, w) \Vdash A$ if $w \Vdash B$ and $n > m$ for all m which appear in w.*

Moreover, if Γ is a sequence of modal formulas, by a witness for Γ, we mean a sequence of witnesses such that any witness w_i in the sequence is a witness for A_i in Γ.

Informally, a witness for a formula A is a sequence of numbers which we assign to occurrences of the boxes in A such that the number for outer box is greater than all numbers of inner boxes. This condition formalizes the idea that any outer box refers to the meta-theories in the hierarchy.

Example 1.3.6. *For instance, $w = (n, m, k, r)$ is a witness for $\Box(p \rightarrow q) \vee \Box(\neg\Box p \rightarrow \Box q)$ if $m > k, r$.*

The next definition is about evaluating a modal formula by an arithmetical substitution for atoms and a witness for the boxes in the formula.

Definition 1.3.7. *Let w be a witness for A and σ an arithmetical substitution which assigns an arithmetical sentence to a propositional variable. And also let $(M, \{T_n\}_{n=0}^{\infty})$ be a provability model. By $A^\sigma(w)$ we mean an arithmetical sentence which results by substituting the variables by the values of σ and interpreting any box as the provability predicate of T_n if the corresponding number in the witness for this box was n. The interpretation of boolean connectives are themselves. Moreover, if Γ is a sequence of modal formulas A_i, and $w = (w_i)_i$ is its witness, by $\Gamma^\sigma(w)$ we mean the sequence of $A_i^\sigma(w_i)$.*

Example 1.3.8. *For the witness and the formula A of the last example, $A^\sigma(w)$ would be $\mathrm{Pr}_{T_n}(p^\sigma \rightarrow q^\sigma) \vee \mathrm{Pr}_{T_m}(\neg\mathrm{Pr}_{T_k}(p^\sigma) \rightarrow \mathrm{Pr}_{T_r}(q^\sigma))$.*

We are ready to introduce the concept of the satisfiability of a formula in a provability model.

Definition 1.3.9. *A sequent $\Gamma \Rightarrow \Delta$ is true in $(M, \{T_n\}_{n=0}^{\infty})$ when there are sequences of expansions $\bar{\Gamma}$ and $\bar{\Delta}$ of Γ and Δ, respectively, and witnesses u and v for $\bar{\Gamma}$ and $\bar{\Delta}$ respectively such that for any arithmetical substitution σ, $M \models \bar{\Gamma}^\sigma(u) \Rightarrow \bar{\Delta}^\sigma(v)$. Moreover, we say that a sequent $\Gamma \Rightarrow \Delta$ is true in a class of models $\mathcal{C}$, when there are uniform sequences of expansions and witnesses for all models. In a more precise way, we write $\mathcal{C} \vDash \Gamma \Rightarrow \Delta$, if there are sequences of expansion $\bar{\Gamma}$ and $\bar{\Delta}$ and witnesses u and v such that for all arithmetical substitutions σ and all provability models $(M, \{T_n\}_{n=0}^{\infty})$ in $\mathcal{C}$, $M \models \bar{\Gamma}^\sigma(u) \Rightarrow \bar{\Delta}^\sigma(v)$.*

Informally speaking, truth means the existence of expansions and witnesses such that the interpretation of a formula (or sequent) becomes true for all arithmetical substitutions.

Remark 1.3.10. *Note that our definition of satisfiability allows us to use a disjunction of finitely many expansions of the formula instead of the original formula itself. In other words, if we want to show that* $(M, \{T_n\}_{n=0}^{\infty}) \vDash A$*, we could use finitely many expansions* $B_1, B_2, \ldots, B_k$ *for* A *and find a witness for* $\bigvee_{i=1}^{k} B_i$*. The same is true for the sequents.*

Let us illuminate the Definition 1.3.9 with some examples.

Example 1.3.11. *Let* $(\mathbb{N}, \{T_n\}_{n=0}^{\infty})$ *be a pair where* $T_0 = \mathrm{PA}$ *and for any* n*,* $T_{n+1} = T_n + \mathrm{Rfn}(T_n)$*. Based on the definition, this pair is obviously a provability model. We want to show that the sentence* $\Box(\Box A \to A)$ *is true in the model. To do this, we need some expansions of the formula and a witness for them. For the expansions, just use the formula itself, and for a witness, first find a witness for* A *and call it* w*; if* n *is a number greater than all the numbers in* w*, then the sequence* $(n+1, n, w, w)$ *is a witness for* $\Box(\Box A \to A)$*. For any arithmetical substitution* σ*, we have* $\mathbb{N} \vDash \mathrm{Pr}_{T_{n+1}}(\mathrm{Pr}_{T_n}(A^\sigma(w)) \to A^\sigma(w))$ *since the theory* T_{n+1} *can prove reflection for* T_n*. As you can see, the idea of introducing a hierarchy to witness the boxes in modal sentences could kill the effect of Gödel's second incompleteness theorem.*
Let us illuminate the importance of expansions with an example. Consider the sentence $\neg\Box(\neg\Box A \wedge A)$*. We want to show that this sentence is true in the above mentioned provability model. (Note that this formula is provable in* **S4***.) Pick a witness* w *for the sentence* A*, a number* n *greater than all numbers in* w *and the formula itself as its expansion. In this case we need two copies of the sentence, therefore we have to find a witness for* $B = \neg\Box(\neg\Box A \wedge A) \vee \neg\Box(\neg\Box A \wedge A)$*. It is easy to verify that the sequence* $(n+2, n+1, w, w, n+1, n, w, w)$ *is a witness for* B*. For any arithmetical substitution* σ*, we have*

$$\mathbb{N} \vDash \neg\mathrm{Pr}_{n+2}(\neg\mathrm{Pr}_{n+1}(A^\sigma(w)) \wedge A^\sigma(w)) \vee \neg\mathrm{Pr}_{n+1}(\neg\mathrm{Pr}_n(A^\sigma(w)) \wedge A^\sigma(w))$$

Because if we have both

$$\mathrm{Pr}_{n+2}(\neg\mathrm{Pr}_{n+1}(A^\sigma(w)) \wedge A^\sigma(w))$$

and

$$\mathrm{Pr}_{n+1}(\neg\mathrm{Pr}_n(A^\sigma(w)) \wedge A^\sigma(w))$$

then from the first part and the soundness of T_{n+2} *we have* $\neg\mathrm{Pr}_{n+1}(A^\sigma(w))$ *and from the second part and the fact that the provability predicate commutes with* $\wedge$*, we have* $\mathrm{Pr}_{n+1}(A^\sigma(w))$*, which is a contradiction. Therefore, the sentence is true in* $\mathbb{N}$*. It is easy to see that if we want to show the truth of the sentence* $\Box(\neg\Box(\neg\Box A \wedge A))$*, we should use the expansion* $\Box B$ *of the formula. This observation shows the importance of the expansions, but is it possible to avoid them?*

Example 1.3.12. *In this example we want to argue that some sentences do not have a witness in some provability models. Finding these kinds of examples is not hard. It is enough to think of formulas such as* p *or* $\Box p$*. However, what we want*

to show here is finding an example to show the importance of the expansions in the definition. Think of the provability model of the last example and consider the formula $\neg\Box(\neg\Box p \wedge p)$*. We showed that if we use two different copies of the formula, then the disjunction of those different copies have a witness in the provability model. We want to show that if we just use one copy, it is impossible to witness the formula. Assume that* $w = (n, m)$ *is a witness for* $\neg\Box(\neg\Box p \wedge p)$ *in the above mentioned provability model. Then since* w *is a witness, we have* $n > m$*. On the other hand, we know that for any arithmetical substitution, we should have* $\mathbb{N} \vDash \neg\mathrm{Pr}_n(\neg\mathrm{Pr}_m(p^\sigma) \wedge p^\sigma)$*. Use the arithmetical substitution which sends* p *to* $\mathrm{Cons}(T_m)$*. Therefore, we have*

$$\mathbb{N} \vDash \neg\mathrm{Pr}_n(\neg\mathrm{Pr}_m(\mathrm{Cons}(T_m)) \wedge \mathrm{Cons}(T_m))$$

Based on the formalized Gödel's second incompleteness theorem

$$I\Sigma_1 \vdash \mathrm{Cons}(T_m) \to \neg\mathrm{Pr}_m(\mathrm{Cons}(T_m))$$

since $I\Sigma_1 \subseteq T_{m+1}$ *and* $T_{m+1} \vdash \mathrm{Cons}(T_m)$ *we have*

$$T_{m+1} \vdash \neg\mathrm{Pr}_m(\mathrm{Cons}(T_m))$$

hence $\mathbb{N} \vDash \mathrm{Pr}_{m+1}(\neg\mathrm{Pr}_m(\mathrm{Cons}(T_m)))$ *and since* T_{m+1} *has the reflection principle for* T_m*,* $\mathbb{N} \vDash \mathrm{Pr}_{m+1}(\mathrm{Cons}(T_m))$*. Since* $n > m$ *we have*

$$\mathbb{N} \vDash \mathrm{Pr}_n(\neg\mathrm{Pr}_m(\mathrm{Cons}(T_m))) \wedge \mathrm{Pr}_n(\mathrm{Cons}(T_m))$$

which contradicts our assumption. As you can see, our provability interpretation is sensitive to the use of expansions and also to the numbers of copies of expansions. In the following discussion, we will argue that this property is an inherent property of the informal intuition behind modal formulas.

1.3.2 Discussion

One of the complexities of our provability interpretation is the use of expansions and in this discussion, we want to justify its role. But before that, we need some observations. First of all, it seems that if we use the intuitive interpretation of the boxes as the provability predicates of different theories in the hierarchy of theories, meta-theories, meta-meta-theories and so on, the natural provability interpretation will be the following:

A sentence A *is true in a provability model* $(M, \{T_n\}_{n=0}^{\infty})$*, if there is a witness* w *for* A *such that for all* σ*,* $M \models A^\sigma(w)$*.*

Which informally says that if you could witness the boxes in the formula A in the provability model, then it is true. Note that this definition is simpler than ours and does not use any kind of expansions. Let us concentrate on **S4** as the theory for our intuitive provability, and temporarily use the above definition as the definition of the truth. To interpret all axioms of the system **S4** it is easy to see that we need two natural conditions on our model. First of all, T_{n+1} should be powerful enough to prove the reflection of the theory T_n and secondly, all T_ns should be sound with respect to our model M (This is what the nature of

provability in **S4** assumes; think of $\Box(\Box A \to A)$ and $\Box A \to A$, respectively.) The sentence $\neg\Box(\neg\Box A \wedge A)$ is a theorem of **S4** and we expect that it should be true in any model with those two conditions. But in Example 1.3.12 we showed that there is no witness for the sentence and hence, with the definition above, the sentence is not true. The reason is the different roles of an occurrence of a box in a modal formula. To elucidate this point, let us investigate the intuitive proof of the sentence $\neg\Box(\neg\Box A \wedge A)$ in **S4**. The proof is a proof by contradiction. Assume $\Box(\neg\Box A \wedge A)$, then because all theorems are true (axiom **T**), we have $\neg\Box A \wedge A$ and hence $\neg\Box A$. On the other hand, since the provability commutes with the conjunction (a consequence of the axiom **K**), we have $\Box A$, which is a contradiction. Consider the fact that the box in $\neg\Box A$ is inherited from the inner box in $\neg\Box A \wedge A$ and the box in $\Box A$ is inherited from the outer box in $\Box(\neg\Box A \wedge A)$. Therefore, to reach the contradiction, we need these two boxes refer to one layer in the hierarchy of theories which is impossible because the inner one is the theory and the other is the meta-theory and it is impossible to have $T_{n+1} = T_n$, because T_{n+1} should prove the reflection for T_n.

What these considerations show, is actually the fact that one box in **S4** could have different roles. (In the above sentence, the outer box has two different roles, one as the meta-theory of the inner box and the other, as the theory itself.) Therefore, the natural way to interpret these boxes is an approach which captures the different roles of a box at the same time, and this is not possible with the above simplified semantics, because it is obviously based on the assumption that any box has just one role which needs just one witness. Here is where we need expansions. In fact, the intended purpose of the expansions is using different copies of the formula, first to allow several attempts at witnessing a single formula and then to capture different roles of one box at once. (See Example 1.3.11 to find out how this technique works.)

There is another question to ask. Why do we need this kind of iterative expansion method and why is just the simple disjunction of the formula not enough? The answer is that for any fixed role available for one box, it is also possible to have different roles for inner boxes. Therefore, after any box you need a new disjunction. (Think of the sentence $\Box(\neg\Box(\neg\Box A \wedge A))$.) This is just what we call expansions.

As a conclusion for this discussion, let us compare our situation here in modal logic with first order logic. In first order logic, if we have a theorem of the form $\forall x \exists y A(x, y)$ where $A(x, y)$ is quantifier-free and if we want to witness y, Herbrand's theorem gives the answer; we can witness y by terms in our language. However, we know that one term is not enough. The reason is simple. The existentially quantified y could have different values (roles) and these different values (roles) can be captured by a disjunction of sentences $A(x, t(x))$ for some finite possible set of terms $t(x)$. The situation in modal logic is the same. We read boxes as existence of theories and we want to witness them. Since there are different roles for any box, we need a disjunction to capture these different roles. In other words, we could interpret the expansions as a kind of Herbrandization of the modal formulas.

1.4 The Logic K4

Intuitively, the logic **K4** is sound with respect to all kinds of provability interpretations. The reason is very simple. **K4** has two important modal axioms; the axiom **K** which means that the provability predicate is closed under modus ponens, and the axioms **4** which means that the provability of a sentence is also provable. The first axiom is a very easy fact and all strong enough meta-theories can prove it. On the other hand, if our meta-theory is sufficiently strong (Σ_1-complete), the axiom **4** would be also easily proved. Consider the fact that these axioms are not only true but also provable and it justifies the use of the necessitation rule. Hence, **K4** is valid in all provability interpretations. In this section we want to formalize this intuitive argument and show that the logic **K4** is sound and also strongly complete with respect to the class of all provability models.

1.4.1 Soundness

If we denote the class of all provability models by **PrM**, we have:

Theorem 1.4.1. *(Soundness) If* $\Gamma \vdash_{\mathbf{K4}} A$ *then* $\mathbf{PrM} \vDash \Gamma \Rightarrow A$.

Proof. To prove the soundness theorem for **K4**, we will use the cut-free sequent calculus for **K4** i.e. $G(\mathbf{K4})$. To simplify the proof, we use the following conventions: Firstly, if Φ and Ψ are sequences of arithmetical sentences and T is an arithmetical theory, by $T \vdash \Phi \Rightarrow \Psi$, we mean $T \vdash \bigwedge \Phi \rightarrow \bigvee \Psi$. Secondly, without loss of generality, we assume that the main formulas in all of the rules, except the exchange rule, are just the rightmost formulas in the sequent. We just use this assumption for the sake of brevity and clarity of the proof.
We want to prove the following claim by induction on the length of the proof in $G(\mathbf{K4})$.

Claim. If $\Gamma \Rightarrow \Delta$ is provable in $G(\mathbf{K4})$, then there are sequences of expansions $\bar{\Gamma}$ and $\bar{\Delta}$ and witnesses w_1 and w_2 for $\bar{\Gamma}$ and $\bar{\Delta}$ respectively such that for any provability model $(M, \{T_n\}_{n=0}^{\infty})$ and any arithmetical substitution σ, $I\Sigma_1 \vdash \bar{\Gamma}^{\sigma}(w_1) \Rightarrow \bar{\Delta}^{\sigma}(w_2)$.

1. The case of axioms and structural rules. For the axiom $A \Rightarrow A$, it is enough to use A as its expansion in both sides and just an arbitrary witness for A in both sides, again.

For the exchange rule, just use the same expansions and witnesses after the application of the corresponding exchange.

For the weakening rule, if we prove $\Gamma, A \Rightarrow \Delta$ from $\Gamma \Rightarrow \Delta$, by IH, we could find expansions $\bar{\Gamma}$, $\bar{\Delta}$ and witnesses w_1 and w_2. Pick an arbitrary witness w for A. For $\Gamma, A \Rightarrow \Delta$, use the sequences $\bar{\Gamma}, A$ and $\bar{\Delta}$, and for the witnesses use (w_1, w) and w_2. It is easy to show that $I\Sigma_1 \vdash \bar{\Gamma}^{\sigma}(w_1), A^{\sigma}(w) \Rightarrow \bar{\Delta}^{\sigma}(w_2)$. The case for the right weakening is the same.

For the contraction rule, if we prove $\Gamma, A \Rightarrow \Delta$ from $\Gamma, A, A \Rightarrow \Delta$, then by IH, there are sequences of expansions $\{\bar{\Gamma}, \{\bar{A}_{i1}\}_{i=0}^{r}, \{\bar{A}_{j2}\}_{j=0}^{s}\}$ and Δ and also witnesses $w_1 = (u, (v_{i1})_{i=0}^{r}, (v_{j2})_{j=0}^{s})$ and w_2. For the sequent $\Gamma, A \Rightarrow \Delta$, use the sequences of expansions $\{\bar{\Gamma}, \{\bar{A}_{i1}\}_{i=0}^{r}, \{\bar{A}_{j2}\}_{j=0}^{s}\}$ and $\bar{\Delta}$ and for the witnesses just use the same witnesses. In this case, because of the use of a finite set of different expansions instead of just one expansion, we can say that the semantics absorbs the contraction rule. The case for the right contraction is the same.

2. The case of propositional rules. In this case we just prove the case that the last rule is $R\wedge$; the other rules are similar and the argument is the same. If $\Gamma_1, \Gamma_2 \Rightarrow \Delta_1, \Delta_2, A \wedge B$, is proved from $\Gamma_1 \Rightarrow \Delta_1, A$ and $\Gamma_2 \Rightarrow \Delta_2, B$ then by IH we have the sequences of expansions $\bar{\Gamma}_1$, $\{\bar{\Delta}_1, \{\bar{A}_i\}_{i=0}^{r}\}$, $\bar{\Gamma}_2$, $\{\bar{\Delta}_2, \{\bar{B}_j\}_{j=0}^{s}\}$ and witnesses w_1 and $w_2 = (u, (x_i)_{i=0}^{r})$ and w_1', $w_2' = (u', (y_j)_{j=0}^{s})$. For the sequent $\Gamma_1, \Gamma_2 \Rightarrow \Delta_1, \Delta_2, A \wedge B$ use the sequences of expansions $\{\bar{\Gamma}_1, \bar{\Gamma}_2\}$, $\{\bar{\Delta}_1, \bar{\Delta}_2, \{\bar{A}_i \wedge \bar{B}_j\}_{i=0,j=0}^{i=r,j=s}\}$ and witnesses (w_1, w_1'), $(u, u', ((x_i, y_j))_{i=0,j=0}^{i=r,j=s})$.

3. The case of modal rules. If $\Box\Gamma \Rightarrow \Box A$ is proved from $\Gamma, \Box\Gamma \Rightarrow A$, then by IH, we have the sequences of expansions $\{\bar{\Gamma}_1, \overline{\Box\Gamma}_2\}$ and $\{A_i\}_{i=0}^{r}$ and witnesses $w_1 = ((u_j)_{j=0}^{s}, (v_k)_{k=0}^{t})$ and $w_2 = (x_i)_{i=0}^{r}$ where u_j is a witness for the jth formula in $\bar{\Gamma}_1$ and v_k is a witness for the kth formula in $\overline{\Box\Gamma}_2$. Pick number n greater than all the numbers in w_1 and w_2. For the sequent $\Box\Gamma \Rightarrow \Box A$ use the sequences of expansions $\{\overline{\Box\Gamma}_1, \overline{\Box\Gamma}_2\}$ and $\Box \bigvee_{i=0}^{r} A_i$ and for the witnesses use $((n, u_j)_{j=0}^{s}, (v_k)_{k=0}^{t})$ and $(n, (x_i)_{i=0}^{r})$. By IH, we know that for any arithmetical substitution σ,

$$I\Sigma_1 \vdash \bigwedge_{j=0}^{s} \bar{\Gamma}_1^{\sigma}(u_j) \wedge \bigwedge_{k=0}^{t} \overline{\Box\Gamma}_2^{\sigma}(v_k) \to \bigvee_{i=0}^{r} A_i^{\sigma}(x_i).$$

Since $I\Sigma_1 \subseteq T_n$, we have

$$T_n \vdash \bigwedge_{j=0}^{s} \bar{\Gamma}_1^{\sigma}(u_j) \wedge \bigwedge_{k=0}^{t} \overline{\Box\Gamma}_2^{\sigma}(v_k) \to \bigvee_{i=0}^{r} A_i^{\sigma}(x_i).$$

Therefore, by Σ_1-completeness in $I\Sigma_1$ we have

$$I\Sigma_1 \vdash \mathrm{Pr}_n(\bigwedge_{j=0}^{s} (\bar{\Gamma}_1^{\sigma}(u_j) \wedge \bigwedge_{k=0}^{t} (\overline{\Box\Gamma}_2^{\sigma}(v_k))) \to \bigvee_{i=0}^{r} A_i^{\sigma}(x_i)),$$

hence

$$I\Sigma_1 \vdash \mathrm{Pr}_n(\bigwedge_{j=0}^{s} \bar{\Gamma}_1^{\sigma}(u_j)) \wedge \mathrm{Pr}_n(\bigwedge_{k=0}^{t} \overline{\Box\Gamma}_2^{\sigma}(v_k)) \to \mathrm{Pr}_n(\bigvee_{i=0}^{r} A_i^{\sigma}(x_i)).$$

By formalized Σ_1-completeness of T_n in $I\Sigma_1$ we have

$$I\Sigma_1 \vdash \bigwedge_{k=0}^{t} \overline{\Box\Gamma}_2^{\sigma}(v_k) \to \mathrm{Pr}_n(\bigwedge_{k=0}^{t} \overline{\Box\Gamma}_2^{\sigma}(v_k))$$

and hence

$$I\Sigma_1 \vdash \bigwedge_{j=0}^{s} \mathrm{Pr}_n(\bar{\Gamma}_1^{\sigma}(u_j)) \wedge \bigwedge_{k=0}^{t} \overline{\Box\Gamma}_2^{\sigma}(v_k) \to \mathrm{Pr}_n(\bigvee_{i=0}^{r} A_i^{\sigma}(x_i)),$$

which is what we wanted to prove and this completes the proof of the claim. □

For the proof of the soundness theorem, if $\Gamma \vdash_{\mathbf{K4}} A$ then there exists a finite set $\Delta \subseteq \Gamma$ such that $\Delta \vdash_{\mathbf{K4}} A$. Therefore, $G(\mathbf{K4}) \vdash \Delta \Rightarrow A$. By the Claim there are some expansions $\bar{\Delta}$ and $\{A_i\}_{i=0}^r$ for Δ and A, respectively and witnesses u and $\{w_i\}_{i=0}^r$ such that for any arithmetical substitution σ, we have $I\Sigma_1 \vdash \bar{\Delta}^\sigma(u) \Rightarrow \bigvee_{i=0}^r A_i^\sigma(w_i)$. Since $M \vDash I\Sigma_1$, we have $M \vDash \bar{\Delta}^\sigma(u) \Rightarrow \bigvee_{i=0}^r A_i^\sigma(w_i)$. Pick $\bar{\Gamma}$ the same as Γ after replacing the part of Δ by $\bar{\Delta}$. Moreover, choose the witness v for $\bar{\Gamma}$ as an arbitrary expansion of u to $\bar{\Gamma}$. Hence, $M \vDash \bar{\Gamma}^\sigma(v) \Rightarrow \bigvee_{i=0}^r A_i^\sigma(w_i)$ which completes the proof of the soundness. □

1.4.2 Completeness

For the completeness theorem, the idea is to reduce the completeness of **K4** to the completeness of **GL** which is the well-known Solovay's theorem. (See Preliminaries and [20].) To do that, we need a translation from **K4** to **GL** which could transfer the provability behavior of **K4** to the provability behavior of **GL**.

Definition 1.4.2. *Let A be a modal formula with k occurrences of $\Box$ and let $Q = \{q_i\}_{i=0}^\infty$ be a sequence of atoms which are not used in A. Then, a translation t based on Q for the modal sentence A, is a sequence of k numbers which assigns natural numbers to boxes in A such that the number assigned to the outer box is greater than all the numbers for the inner boxes. And A^t is defined as follows:*

(i) *If A is an atom, $A^t = A$.*

(ii) $(B \circ C)^t = B^t \circ C^t$ *for all* $\circ \in \{\wedge, \vee, \rightarrow\}$

(iii) $(\neg B)^t = \neg B^t$.

(iv) $(\Box B)^t = \Box(\bigwedge_{i=0}^n q_i \rightarrow B^t)$ *where n is the number assigned to the box in t.*

Informally, if we interpret a box as the provability predicate for the theory S, then the translation t is just changing the provability predicate of the theory S to the provability predicate of the theory $S + \{q_0, \ldots, q_n\}$ where n is the number that t assigns to that box. For instance, if $t = (1, 2, 1)$ and $A = \Box p \rightarrow \Box\Box p$, then A^t will be the following modal formula:

$$\Box(q_0 \wedge q_1 \rightarrow p) \rightarrow \Box(q_0 \wedge q_1 \wedge q_2 \rightarrow \Box(q_0 \wedge q_1 \rightarrow p)).$$

We want to show that this translation is complete, i.e.

Theorem 1.4.3. *If* $\mathbf{GL} \vdash A^t$ *for some translation t, then* $\mathbf{K4} \vdash A$.

The natural proof should be based on a technique of the transformation of transitive Kripke models to conversely well-founded transitive Kripke models, which is implemented by the following lemma.

Lemma 1.4.4. *Let (K, R, V) be a finite transitive Kripke tree with clusters, A a modal formula and t a translation. Then there is a finite transitive irreflexive Kripke model (K', R', V') such that for any node $k \in K$, there is a node $k' \in K'$ such that if $k \vDash A$ then $k' \vDash A^t$.*

Proof. First of all, for all subformulas B of A, define the complexity of B, $C(B)$, as follows: If B is box-free, define $C(B) = -1$. Otherwise, define $C(B)$ as the maximum number assigned by t in B. Moreover, suppose that $C(A) = n$. To simplify the proof, let us make some conventions. We will use I for clusters and for any $k \in K$, by $I(k)$ we mean the cluster of k. By a path $p = (k_\alpha)_{\alpha=0}^{M}$, we mean a sequence of nodes in K such that for any α, $(k_\alpha, k_{\alpha+1}) \in R$ and if all the nodes of the path p belong to the cluster I, we write $p \subset I$. Moreover, we write $p \prec p'$, when p is a proper initial segment of p'. Finally, by $e(p)$ we mean the rightmost element of p, or in other words, the end point of p.

For any cluster I define $X(I)$ as follows: If I consists of one irreflexive node k, $X(I) = \{k\}$ and if I consists of reflexive nodes, define $X(I)$ as the subset of all paths $p \subset I$ with length less than or equal to $n+2$. The idea is simple. We want to transform a transitive model to a nonreflexive transitive model. To accomplish this, we will unwind the reflexive clusters by some paths of nodes in that cluster and we will use variables in Q to refer to a copy of the node instead of itself, when we check the truth of the modal formulas.

Define $K' = \bigcup_I X(I)$ and $R' = R_1 \cup R_2$ where

$$R_1 = \bigcup_{(k,l)\in R, I(k)\neq I(l)} \{(a,b) \mid a \in X(I(k)) \text{ and } b \in X(I(l))\}$$

and

$$R_2 = \bigcup_I \{(p,p') \mid p \prec p'; p, p' \subset I\}.$$

And finally, define

$$V'(r) = \{p \in K' \mid e(p) \in V(r)\} \cup \{k \mid k \in V(r) \text{ and } k \text{ is irreflexive}\}$$

for all atoms r in A, and

$$V(q_i) = \{k \mid k \text{ is irreflexive}\} \cup \{p \mid |p| \leq n+2-i\}.$$

Informally speaking, K' is just the set K where you replace each reflexive cluster I with all paths of length less that or equal to $n+2$ of nodes in I; R' and V' are the natural relation and valuation induced by R and V, respectively and q_i is true in all irreflexive nodes and also in all paths of nodes in reflexive clusters with length bounded by $n+2-i$. We want to prove the following two claims.

Claim.1. The model (K', R', V') is a finite transitive irreflexive Kripke model.

The finiteness follows from the definition. For the transitivity, suppose that $a, b, c \in K'$ and $(a,b) \in R'$ and $(b,c) \in R'$. Then, there are two cases. The first case is when a and b come from the same cluster. Hence, by definition, this cluster should be a reflexive cluster. Therefore, a and b are paths in this cluster and $a \prec b$. If c comes also from this cluster, we will have $b \prec c$ and since $\prec$ is transitive, we have $a \prec c$ and hence $(a,c) \in R'$. But, if c comes from another cluster, then the cluster of c should be above the cluster of b and hence it is also above the cluster of a which is the same as b's and then by definition we have

$(a, c) \in R'$.
The proof of the second case, which is when a and b come from different clusters, is similar to the proof of the first case.

For the irreflexivity, suppose $(a, a) \in R'$. If a is an irreflexive node in K, then it is impossible, by the definition of R', to have $(a, a) \in R'$. If a comes from a reflexive cluster, then again by the definition of R', the path a should be a proper segment of itself which is impossible.

Claim.2. For all subformulas B of A, if $k \vDash B$, then

$$\begin{cases} \forall p, |p| \leq n + 1 - C(B) \wedge e(p) = k,\ p \vDash B^t & \text{if } k \text{ is reflexive.} \\ k \vDash B^t & \text{if } k \text{ is irreflexive.} \end{cases}$$

and if $k \nvDash B$ then

$$\begin{cases} \forall p, |p| \leq n + 1 - C(B) \wedge e(p) = k,\ p \nvDash B^t & \text{if } k \text{ is reflexive.} \\ k \nvDash B^t & \text{if } k \text{ is irreflexive.} \end{cases}$$

To prove the claim, we use induction on B.

1. Atomic case. If B is an atom, the claim easily follows from the definition of V'.

2. If $B = C \wedge D$ and $k \vDash C \wedge D$ then $k \vDash C$ and $k \vDash D$. If k is irreflexive, then by IH, the claim holds. If k is reflexive, then by IH, for all p such that $|p| \leq n + 1 - C(C)$ and $e(p) = k$, we have $p \vDash C^t$. And also for all p such that $|p| \leq n + 1 - C(D)$ and $e(p) = k$, we have $p \vDash D^t$, and since $C(C \wedge D) = max\{C(C), C(D)\}$, then for all p such that $|p| \leq n + 1 - C(C \wedge D)$ and $e(p) = k$, we have $p \vDash C^t \wedge D^t$.
If $k \nvDash C \wedge D$, then $k \nvDash C$ or $k \nvDash D$. W.l.o.g. assume $k \nvDash C$. If k is irreflexive, the claim is obvious. If k is reflexive, then by IH, for all p such that $|p| \leq n+1-C(B)$ and $e(p) = k$ we have $p \nvDash C^t$, and again since $C(C \wedge D) = max\{C(C), C(D)\}$ we have $\forall p, |p| \leq n + 1 - C(B \wedge D) \wedge e(p) = k,\ p \nvDash (C \wedge D)^t$.

3. If $B = \neg C$, then for irreflexive k, the claim is obvious from IH. If k is reflexive and $k \vDash \neg C$, then $k \nvDash C$, and by IH, $\forall p, |p| \leq n + 1 - C(C)\ p \nvDash C^t$. Therefore, $\forall p, |p| \leq n + 1 - C(C)\ p \vDash \neg C^t$ and since $C(C) = C(\neg C)$ we have what we wanted. The other case is the dual of the first case.

4. The case for disjunction and implication is the same as the cases for conjunction and negation and we omit them here.

5. The modal case. This is the most important and the most complex part of the proof.

5.1. If $B = \Box C$ and $k \vDash \Box C$ then for all l which $(k, l) \in R$, $l \vDash C$. Define $m = C(B)$.
5.1.1. If k is irreflexive, we know that the nodes above k in K' are of two forms.

The l's which are irreflexive and $(k,l) \in R$ or the p's where p comes from a cluster I above k and $e(p) = l$. For the first kind of nodes, by IH we know that $l \vDash C^t$, therefore $l \vDash \bigwedge_{i=0}^{m} q_i \to C^t$. If we were in the second case, we know that $l \vDash C$ and again by IH, for all p such that $|p| \leq n+1-C(C)$ and $e(p) = l$, we have $p \vDash C^t$. Therefore, for all $p, |p| \leq n+1-C(C)$ we have $p \vDash C^t$ and hence $p \vDash \bigwedge_{i=0}^{k} q_i \to C^t$. If $|p| > n+1-C(C)$, since $C(C) < C(B) = m$, we have $|p| > n+2-m$, and then by the definition of the valuation we know that $p \nvDash q_m$ and hence $p \nvDash \bigwedge_{i=0}^{m} q_i$ and thus $p \vDash \bigwedge_{i=0}^{m} q_i \to C^t$. Therefore, for all p above k, we have $p \vDash \bigwedge_{i=0}^{m} q_i \to C^t$. Since for all nodes above k, $\bigwedge_{i=0}^{m} q_i \to C^t$ is true, we have $k \vDash \Box(\bigwedge_{i=0}^{m} q_i \to C^t)$ which means $k \vDash (\Box C)^t$.

5.1.2. If k is reflexive from the cluster I, pick p such that $|p| \leq n+1-m$. We want to show that $p \vDash \Box(\bigwedge_{i=0}^{m} q_i \to C^t)$. We know that all nodes above p are of the form irreflexive l's or $p' \subset J$ where J is a cluster above I or $p' \subset I$ where $p \prec p'$. For the first and second kinds, by a proof similar to that of 5.1.1, we can show that $l \vDash \bigwedge_{i=0}^{m} q_i \to C^t$ and $p' \vDash \bigwedge_{i=0}^{m} q_i \to C^t$. For the third case, if $|p'| > n+2-m$, then $p' \nvDash q_m$ and hence $p' \nvDash \bigwedge_{i=0}^{m} q_i$ and thus $p' \vDash \bigwedge_{i=0}^{m} q_i \to C^t$. If $|p'| \leq n+2-m$ then since $C(C) \leq m-1$ we have $|p'| \leq n+1-C(C)$. On the other hand, $k \vDash \Box C$, hence all nodes in I satisfy C, and specially we have $e(p') \vDash C$, by IH, and by the fact that $|p'| \leq n+1-C(C)$, we have $p' \vDash C^t$ and therefore $\bigwedge_{i=0}^{m} q_i \to C^t$. We proved that at all nodes above $p \in K'$, we have $\bigwedge_{i=0}^{m} q_i \to C^t$ hence $p \vDash \Box(\bigwedge_{i=0}^{m} q_i \to C^t)$ which is what we wanted.

5.2. If $B = \Box C$ and $k \nvDash \Box C$, then there is a node l such that $l \nvDash C$. Define $C(B) = m$.
5.2.1. If k is irreflexive, we want to show that $k \nvDash \Box(\bigwedge_{i=0}^{m} q_i \to C^t)$. Note that since $(k,l) \in R$, and k is irreflexive, then $l \neq k$ and it belongs to a cluster above k. If l is irrefelexive then by IH, $l \nvDash C^t$ and also since it is irreflexive, for all i, $l \vDash q_i$; hence $l \nvDash \bigwedge_{i=0}^{m} q_i \to C^t$ since $l \neq k$ and $(k,l) \in R$, $(k,l) \in R'$. Therefore, $k \nvDash \Box(\bigwedge_{i=0}^{m} q_i \to C^t)$. If l is a reflexive node of the cluster I, then define $p \subset I$ as a path such that $|p| = n+2-m$ and $e(p) = l$. Since $C(C) \leq m-1$ then $|p| \leq n+1-C(C)$. By IH, $p \nvDash C^t$. (Consider that m is the complexity of a boxed formula and therefore $m \geq 0$, hence $n+2-m \leq n+2$ and it means such a p exists.). Moreover, we know that $p \vDash \bigwedge_{i=0}^{m} q_i$ since $|p| \leq n+2-i$ for all $i \leq m$, therefore, $p \nvDash \bigwedge_{i=0}^{m} q_i \to C^t$. Since the cluster of k and the cluster of l are different and $(k,l) \in R$, then $(k,p) \in R'$ and it means that $k \nvDash \Box(\bigwedge_{i=0}^{m} q_i \to C^t)$.

5.2.2. Consider the case that k is reflexive. In this case, if l belongs to a cluster above k, then the proof is the same as 5.2.1. If the cluster of l and k are the same (say I), we have the following construction: Pick p such that $e(p) = k$ and $|p| \leq n+1-m$. We want to show that $p \nVdash \Box(\bigwedge_{i=0}^{m} q_i \to C^t)$. Pick $p' \subset I$ such that $e(p') = l$, $p \prec p'$ and $|p'| = n+2-m$. (It is enough to extend p to a path of length $n+2-m$ ending at l. Note that $n+2-m > n+1-m$, which guarantee the existence of an expansion with endpoint l possibly different from k. Moreover, this length is less that $n+2$ and therefore p' exists in our model as a path). We know that $C(C) \leq m-1$, hence $|p'| \leq n+1-C(C)$. By IH, $p' \nvDash C^t$. On the other hand, $p \vDash \bigwedge_{i=0}^{m} q_i$ since $|p| \leq n+2-i$ for all $i \leq m$, therefore, $p \nvDash \bigwedge_{i=0}^{m} q_i \to C^t$. Since $p \prec p'$, we can conclude that $p \nvDash \Box(\bigwedge_{i=0}^{m} q_i \to C^t)$.

The lemmas are obvious by the claim 2. For B in the claim, choose A itself, then if $k \vDash A$ and k is irreflexive , then $k \vDash A^t$. But if k is reflexive, pick $p = k$ as a path with length one. Hence $|p| = 1 \leq n + 1 - C(A)$, since $C(A) = n$ and therefore, $p \vDash A^t$. Therefore, for any $k \vDash A$ there is a node $k' \in K'$ such that $k' \vDash A^t$. □

For the proof of Theorem 1.4.3 we have:

Proof. If $\mathbf{K4} \nvdash A$, then there is a finite transitive Kripke tree with clusters (K, R, V) and a node k such that $k \vDash \neg A$. If we apply Lemma 1.4.4 for $\neg A$, we can construct a finite transitive irreflexive Kripke model (K', R', V') and a node k' such that $k' \nvDash \neg A^t$. But (K', R', V') is a model of $\mathbf{GL}$ and $\mathbf{GL} \vdash A^t$. A contradiction. Hence $\mathbf{K4} \vdash A$. □

Based on the completeness of the translations, which we have introduced, we are able to prove the completeness theorem. But, since we want to establish a more powerful completeness result, i.e. the strong completeness, we need one more lemma.

Lemma 1.4.5. *There is a hierarchy of theories $\{T_n\}_{n=0}^{\infty}$ such that for any n, $I\Sigma_1 \subseteq T_n$ and $T_n \subseteq T_{n+1}$ provably in $I\Sigma_1$ and also an arithmetical substitution $*$ such that for any modal formula A, if there exists a witness w for A such that $(M, \{T_n\}_{n=0}^{\infty}) \vDash A^*(w)$ for all $M \vDash I\Sigma_1$, then $\mathbf{K4} \vdash A$.*

Proof. Add infinitely many new atoms $Q = \{q_n\}_{n=0}^{\infty}$ to the language of modal logics, and apply all axioms and rules of the logic $\mathbf{K4}$ to the new language to construct a new system $\mathbf{K4}(\mathbf{Q})$ and do the same thing for the logic $\mathbf{GL}$ to construct $\mathbf{GL}(\mathbf{Q})$. Pick the substitution $*$ as the uniform substitution of Solovay's theorem (see Preliminaries and [7]). It simply says that for any A, $I\Sigma_1 \vdash A^*$ iff $\mathbf{GL}(\mathbf{Q}) \vdash A$, where A^* means the combination of substituting any atom p with p^* and interpreting all boxes as the provability predicate of $I\Sigma_1$. For any n, define $T_n = I\Sigma_1 + \{q_i^*\}_{i=0}^{n}$. We claim that this $*$ and this hierarchy $\{T_n\}_{n=0}^{\infty}$ works for the claim of the lemma. First of all, it is easy to show that the hierarchy satisfied the conditions claimed. Secondly, we have $M \vDash A^*(w)$ for all $M \vDash I\Sigma_1$. Therefore, $I\Sigma_1 \vdash A^*(w)$. Use q_i's in the translations from $\mathbf{K4}$ to $\mathbf{GL}$. Since the interpretation of a box in any formula $\Box D$ with witness m is $\mathrm{Pr}_{T_m}(D)$, and it is provably equivalent to $\mathrm{Pr}_{I\Sigma_1}(\bigwedge_{i=0}^{m} q_i \to D)$, it is easy to see that there is a translation t, such that $I\Sigma_1 \vdash A^*(w) \leftrightarrow (A^t)^*$. (In fact t equals to the witness w.) Therefore, $I\Sigma_1 \vdash (A^t)^*$, by the uniform version of Solovay's theorem, $\mathbf{GL}(\mathbf{Q}) \vdash A^t$, and by Theorem 1.4.3, $\mathbf{K4}(\mathbf{Q}) \vdash A$. It means that there exists a proof for A in $\mathbf{K4}(\mathbf{Q})$. Since A does not have any $q_i \in Q$, it is enough to put $q_i = \top$ everywhere in the proof to find a proof for A in $\mathbf{K4}$. □

We want to prove the strong completeness theorem.

Theorem 1.4.6. *(Strong Completeness) If $\mathbf{PrM} \vDash \Gamma \Rightarrow A$, then $\Gamma \vdash_{\mathbf{K4}} A$.*

Proof. We know that there are the sequence of expansions $\bar{\Gamma}$, and expansions $B_1, \ldots, B_k$ of A and witnesses u for $\bar{\Gamma}$, and $w_1, \ldots, w_k$ for $B_1, \ldots, B_k$ such that for all provability models and all arithmetical substitutions σ,

$$M \vDash \bar{\Gamma}^{\sigma}(u) \Rightarrow \{B_i^{\sigma}(w_i)\}_{i=0}^{k}.$$

Pick the hierarchy of theories and $*$ from Lemma 1.4.5. Then for all $M \vDash I\Sigma_1$,

$$M \vDash \bar{\Gamma}^*(u) \Rightarrow \{B_i^*(w_i)\}_{i=0}^k.$$

Hence

$$I\Sigma_1 + \bar{\Gamma}^*(u) \vdash \bigvee_{i=0}^k B_i^*(w_i).$$

Therefore there is a finite $\Delta \subseteq \bar{\Gamma}$ and a subset of witnesses v from u, such that

$$I\Sigma_1 + \Delta^*(v) \vdash \bigvee_{i=0}^k B_i^*(w_i).$$

Hence, for all $M \vDash I\Sigma_1$, we have

$$M \vDash \bigwedge \Delta^*(v) \to \bigvee_{i=0}^k B_i^*(w_i).$$

By Lemma 1.4.5, $\mathbf{K4} \vdash \bigwedge \Delta \to \bigvee_{i=0}^k B_i$, which means $\bar{\Gamma} \vdash_{\mathbf{K4}} \bigvee_{i=0}^k B_i$. Finally, since in the presence of the axiom $\mathbf{K}$, all expansions of a formula are equivalent to itself, $\Gamma \vdash_{\mathbf{K4}} A$. □

1.5 The Logic KD4

The logic **KD4** is a modal logic resulting from adding the axiom $\mathbf{D} : \Box A \to \neg\Box\neg A$ or equivalently $\neg\Box\bot$ to **K4**. Therefore, intuitively, if, in M, all the theories from the hierarchy are consistent and each theory proves the consistency of preceding theories, then the axioms of **KD4** should be valid. (Since we have the neccesitation rule, the sentence $\Box\neg\Box\bot$ is also provable and this is why we need the consistency statements to be provable, as well.) The formalization of these models is exactly what we will call consistent provability models and we will show that the logic **KD4** is sound and strongly complete with respect to these models.

Definition 1.5.1. *A provability model $(M, \{T_n\}_{n=0}^\infty)$ is called consistent if for all n, M thinks that T_n is consistent and $T_{n+1} \vdash \mathrm{Cons}(T_n)$, i.e. $M \vDash \mathrm{Cons}(T_n)$ and $M \vDash \mathrm{Pr}_{T_{n+1}}(\mathrm{Cons}(T_n))$. Moreover, the class of all consistent provability models will be denoted by* **Cons**.

Let us prove the soundness theorem.

Theorem 1.5.2. *(Soundness) If $\Gamma \vdash_{\mathbf{KD4}} A$, then $\mathbf{Cons} \vDash \Gamma \Rightarrow A$.*

Proof. We use the soundness theorem for **K4**. If $\Gamma \vdash_{\mathbf{KD4}} A$, then

$$\Gamma + \Box\neg\Box\bot \wedge \neg\Box\bot \vdash_{\mathbf{K4}} A.$$

Based on the soundness of **K4**, there are sequences $\bar{\Gamma} + \{\Box(\bigvee_{j=0}^{s_i} \neg\Box\bot) \wedge \neg\Box\bot\}_{i\in I}$ and $\{A_k\}_{k=0}^t$ as the expansions of $\Gamma + \Box\neg\Box\bot \wedge \neg\Box\bot$ and A, respectively and

witnesses u, $(n_i, (m_{ij})_{j=0}^{s_i}, k_i)$ and w_k such that for any provability model like $(M, \{T_n\}_{n=0}^{\infty})$ and any arithmetical substitution σ,

$$M \vDash \bar{\Gamma}^{\sigma}(u) + \{\mathrm{Pr}_{n_i}(\bigvee_{j=0}^{s_i} \neg \mathrm{Pr}_{m_{ij}}(\bot)) \wedge \neg \mathrm{Pr}_{k_i}(\bot)\}_{i \in I} \Rightarrow \bigvee_{k=0}^{t} A_k^{\sigma}(w_k))$$

If we apply this fact to the consistent provability models, since $n_i > m_{ij}$ and for any n, $M \vDash \mathrm{Pr}_{n+1}(\neg \mathrm{Pr}_n(\bot))$, we have $M \vDash \mathrm{Pr}_{n_i}(\neg \mathrm{Pr}_{m_{ij}}(\bot))$ for all $i \leq r$ and $j \leq s_i$. Moreover, since for any n, $M \vDash \neg \mathrm{Pr}_n(\bot)$, we have $M \vDash \neg \mathrm{Pr}_{k_i}(\bot)$. Therefore, for any consistent provability model $(M, \{T_n\}_{n=0}^{\infty})$ we have

$$M \vDash \bar{\Gamma}^{\sigma}(u) \Rightarrow \bigvee_{k=0}^{t} A_k^{\sigma}(w_k)$$

which completes the proof of the soundness for **KD4**. □

For the completeness theorem, the idea is reducing the completeness of **KD4** to the completeness of **K4** which was proved in the previous section.

Theorem 1.5.3. *(Strong Completeness) If* $\mathbf{Cons} \vDash \Gamma \Rightarrow A$, *then* $\Gamma \vdash_{\mathbf{KD4}} A$.

Proof. We know that there are a multiset $\bar{\Gamma}$, and expansions $B_1, \ldots, B_k$ of A and witnesses u for $\bar{\Gamma}$, and $w_1, \ldots, w_k$ for $B_1, \ldots, B_k$ such that for any consistent provability model and any arithmetical substitution σ,

$$(M, \{T_n\}_{n=0}^{\infty}) \vDash \bar{\Gamma}^{\sigma}(u) \Rightarrow \{B_i^{\sigma}(w_i)\}_{i=0}^{k}.$$

Define Δ as a sequence which consists of an infinite number of the formula $\Box\neg\Box\bot$ and also an infinite number of the formula $\neg\Box\bot$. We claim that $\Gamma, \Delta \Rightarrow A$ is true in the class **PrM**. For the expansions, use the same expansions for Γ and A, and also use Δ itself, as its sequence of expansions. For witnesses, use u, w_i's and for Δ, for any number n, use $(n+1, n)$ for one of the formulas $\Box\neg\Box\bot$ and n for one of the formulas $\neg\Box\bot$. Call this witness v. Let $(M, \{T_n\}_{n=0}^{\infty})$ be an arbitrary provability model. We claim that

$$M \vDash \bar{\Gamma}^{\sigma}(u), \Delta^{\sigma}(v) \Rightarrow \{B_i^{\sigma}(w_i)\}_{i=0}^{k}.$$

Because when $M \vDash \bar{\Gamma}^{\sigma}(u), \Delta^{\sigma}(v)$ then $M \vDash \Delta^{\sigma}(v)$ which means for any n,

$$M \vDash \mathrm{Pr}_{n+1}(\neg \mathrm{Pr}_n(\bot)),$$

and

$$M \vDash \neg \mathrm{Pr}_n(\bot).$$

Therefore, $(M, \{T_n\}_{n=0}^{\infty})$ is a consistent provability model and since $M \vDash \bar{\Gamma}^{\sigma}(u)$ we have,

$$(M, \{T_n\}_{n=0}^{\infty}) \vDash \bigvee_{i=0}^{k} B_i^{\sigma}(w_i).$$

Therefore, for all provability models and all σ, we have

$$M \vDash \bar{\Gamma}^{\sigma}(u), \Delta^{\sigma}(v) \Rightarrow \{B_i^{\sigma}(w_i)\}_{i=0}^{k}.$$

Hence, by the strong completeness of **K4**, we have $\Gamma, \Delta \vdash_{\mathbf{K4}} A$ and since all formulas in Δ are provable in **KD4**, we have $\Gamma \vdash_{\mathbf{KD4}} A$. □

Remark 1.5.4. *Note that the truth of a formula in a class of provability models means the existence of a uniform sequence of expansions and also a uniform witness for it. In other words, we have a fixed sequence of natural numbers which works for all provability models in the class. Therefore, we could claim that sentences just describe the behavior of the natural numbers instead of some actual theories. What does it mean? It means that sentences do not describe the behavior of a concrete specific provability model, but instead, they talk about the roles of these ingredients in the structure (provability model) which are encoded by the natural numbers. Informally speaking, sentences just transcend the actual theories to their abstract roles in the structure of a provability model. (As an example, think of how the cardinal numbers transcend the concept of cardinality from the actual sets.) For instance, in the case of the logic* **KD4***, it describes the relation between a meta-theory T_{n+1} and its theory T_n which is the condition that the meta-theory is powerful enough to show the consistency of the theory. This is not about some actual theories which we use; it is about the power of the meta-theory in comparison to its theory. In other words,* **KD4** *describes the abstract condition of consistency and provability of consistency. This fact is true in all soundness-completeness results we propose in this chapter.*

1.6 The Logic S4

Intuitively, if we have the property that all theories are sound and the soundness of theories are also provable in their meta-theories, all axioms of **S4**, would be valid. The formalization of these models is exactly what we will call the reflexive provability models. In fact, we will show that the logic **S4** is sound and also strongly complete with respect to the class of all reflexive provability models.

1.6.1 Soundness

First of all we need a definition:

Definition 1.6.1. *A provability model $(M, \{T_n\}_{n=0}^{\infty})$ is reflexive if for any n, M thinks that T_n is sound and $T_{n+1} \vdash \mathrm{Rfn}(T_n)$, i.e. $M \vDash \mathrm{Pr}_{T_n}(A) \to A$ and $M \vDash \mathrm{Pr}_{T_{n+1}}(\mathrm{Pr}_{T_n}(A) \to A)$ for each sentence A. Moreover, the class of all reflexive provability models will be denoted by* **Ref**.

Let us prove the soundness theorem.

Theorem 1.6.2. *(Soundness) If $\Gamma \vdash_{\mathbf{S4}} A$, then* **Ref** $\vDash \Gamma \Rightarrow A$.

Proof. To prove the soundness theorem, we will use the cut-free sequent calculus for **S4**, i.e. $G(\mathbf{S4})$. And, we will use the conventions of Theorem 1.4.1. We want to prove the following claim:

Claim. If $\Gamma \Rightarrow \Delta$ is provable in $G(\mathbf{S4})$, then there are sequences of expansions $\bar{\Gamma}$ and $\bar{\Delta}$ and also witnesses w_1 and w_2 for $\bar{\Gamma}$ and $\bar{\Delta}$, respectively and a number n greater than all the numbers in w_1 and w_2, such that for any reflexive provability model $(M, \{T_n\}_{n=0}^{\infty})$ and any arithmetical substitution σ, $T_n \vdash \bar{\Gamma}^{\sigma}(w_1) \Rightarrow \bar{\Delta}^{\sigma}(w_2)$ is true in M. We will call the number n the context number.

The proof of the claim is by induction on the length of the proof of $\Gamma \Rightarrow \Delta$ and the proof for the non-modal cases are similar to the proof of Theorem 1.4.1. But the difference is just the presence of the context number n here. To find this number in all non-modal cases, if the case is the axiom case, any number works; for contraction and exchange, just use the same number in the induction hypothesis. For weakening, use the successor of the maximum of the context number of the induction hypothesis and the arbitrary chosen witness for the weakening formula. For the other cases, it is enough to use the maximum numbers of the induction hypothesis. We want to prove the case of the modal rules.

1. If $\Gamma, \Box A \Rightarrow \Delta$ is proved by $\Gamma, A \Rightarrow \Delta$, then by IH, we can find sequences of expansions $\{\bar{\Gamma}, \{A_i\}_{i=0}^r\}$, $\bar{\Delta}$ and witnesses $w_1 = (u, (x_i)_{i=0}^r)$ and w_2 and the context number n. For the sequent $\Gamma, \Box A \Rightarrow \Delta$, use the sequences of expansions $\{\bar{\Gamma}, \{\Box A_i\}_{i=0}^r\}$, $\bar{\Delta}$ and for the witnesses use $(u, ((n, x_i))_{i=0}^r)$, w_2 and for the context number use $n+1$. By IH, we know that for all reflexive provability models and all arithmetical substitution σ, M thinks

$$T_n \vdash \bar{\Gamma}^\sigma(w_1), \{A_i^\sigma(x_i)\}_{i=0}^r \Rightarrow \bar{\Delta}^\sigma(w_2).$$

We claim that there is a proof, formalizable in $I\Sigma_1$, for the following statement: If $T_n \subseteq T_{n+1}$, $T_{n+1} \vdash \mathrm{Pr}_n(A_i^\sigma(x_i)) \to A_i^\sigma(x_i)$ for all $i \leq r$ and

$$T_n \vdash \bar{\Gamma}^\sigma(w_1), \{A_i^\sigma(x_i)\}_{i=0}^r \Rightarrow \bar{\Delta}^\sigma(w_2)$$

then

$$T_{n+1} \vdash \bar{\Gamma}^\sigma(w_1), \{\mathrm{Pr}_n(A_i^\sigma(x_i))\}_{i=0}^r \Rightarrow \bar{\Delta}^\sigma(w_2).$$

The proof is simple. We have $T_n \subseteq T_{n+1}$ and $T_{n+1} \vdash \mathrm{Pr}_n(A_i^\sigma(x_i)) \to A_i^\sigma(x_i)$. Therefore,

$$T_{n+1} \vdash \bar{\Gamma}^\sigma(w_1), \{\mathrm{Pr}_n(A_i^\sigma(x_i))\}_{i=0}^r \Rightarrow \bar{\Delta}^\sigma(w_2).$$

The proof just uses the fact that all first order tautologies are provable and Pr is closed under modus ponens and all of these properties are provable in $I\Sigma_1$. Since $M \vDash I\Sigma_1$, M thinks that this implication is true. On the other hand both of premises are true in M, because of IH and the condition of being a reflexive provability model. Therefore, M thinks

$$T_{n+1} \vdash \bar{\Gamma}^\sigma(w_1), \{\mathrm{Pr}_n(A_i^\sigma(x_i))\}_{i=0}^r \Rightarrow \bar{\Delta}^\sigma(w_2),$$

which completes the proof.

2. If $\Box\Gamma \Rightarrow \Box A$ is proved by $\Box\Gamma \Rightarrow A$, then by IH we have sequences of expansions $\overline{\Box\Gamma}$ and some expansions $\{A_i\}_{i=0}^r$ and witnesses w_1 and $(x_i)_{i=0}^r$ and a context number n such that for all arithmetical substitutions σ, M thinks

$$T_n \vdash \overline{\Box\Gamma^\sigma(w_1)} \Rightarrow \{A_i^\sigma(x_i)\}_{i=0}^r.$$

For the sequent $\Box\Gamma \Rightarrow \Box A$, use the expansion $\overline{\Box\Gamma}$ and $\Box(\bigvee_{i=0}^r A_i)$, and the witnesses w_1 and $(n, (x_i)_{i=0}^r)$ and the context number $n+1$.
Based on the Σ_1-completeness available in M, M thinks

$$I\Sigma_1 \vdash \mathrm{Pr}_n(\bigwedge \overline{\Box\Gamma^\sigma(w_1)} \to \bigvee_{i=0}^r (A_i^\sigma(x_i))).$$

Because the provability predicate commutes with the implications provably in $I\Sigma_1$, we have this property in M, hence

$$I\Sigma_1 \vdash \mathrm{Pr}_n(\bigwedge \overline{\Box\Gamma^\sigma(w_1)}) \to \mathrm{Pr}_n(\bigvee_{i=0}^{r}(A_i^\sigma(x_i)))$$

is true in M. Again by Σ_1-completeness, we have

$$I\Sigma_1 \vdash \bigwedge(\overline{\Box\Gamma^\sigma(w_1)}) \to \mathrm{Pr}_n(\bigvee_{i=0}^{r}(A_i^\sigma(x_i)))$$

true in M. And finally since T_{n+1} is an expansion of $I\Sigma_1$ provably in $I\Sigma_1$, we have the inclusion in M, hence

$$T_{n+1} \vdash \bigwedge(\overline{\Box\Gamma^\sigma(w_1)}) \to \mathrm{Pr}_n(\bigvee_{i=0}^{r}(A_i^\sigma(x_i)))$$

is true in M which completes the proof of the claim.
For the proof of the soundness theorem, if $\Gamma \vdash_{\mathbf{S4}} A$ then there exists a finite subset Δ of Γ such that $\Delta \vdash_{\mathbf{S4}} A$. Then $G(\mathbf{S4}) \vdash \Delta \Rightarrow A$, then by the claim, there are sequences of expansions $\bar{\Delta}$ and $\{A_i\}_{i=0}^r$ and the witnesses u and $(x_i)_{i=0}^r$ and a context number n such that for all reflexive provability models $(M, \{T_n\}_{n=0}^\infty)$ and all arithmetical substitution σ, we have $T_n \vdash \bar{\Delta}^\sigma(u) \Rightarrow \bigvee_{i=0}^r A_i^\sigma(x_i)$ in M. Therefore, by soundness of T_n in M, we have $M \vDash \bar{\Delta}^\sigma(u) \Rightarrow \bigvee_{i=0}^r A_i^\sigma(x_i)$. Define $\bar{\Gamma}$ as the sequence of expansions of Γ by using Γ and replacing the subset Δ by $\bar{\Delta}$ and also use any arbitrary witnesses to extend u to a witness for $\bar{\Gamma}$. Call this new witness v. We have

$$M \vDash \bar{\Gamma}^\sigma(v) \Rightarrow \bigvee_{i=0}^{r} A_i^\sigma(x_i)$$

which is what we wanted to prove. □

1.6.2 Completeness

For the completeness theorem, the idea is the same as the idea of the original proof of Solovay's theorem. We will modify the technique of encoding Kripke models in arithmetic. In this case, we need to encode transitive reflexive trees with clusters. Therefore we have two tasks. Firstly, finding a method to encode the clusters and secondly, modifying Solovay's construction to work with reflexive trees instead of irreflexive ones.

Lemma 1.6.3. *Let m be a natural number and $\{T_n\}_{n=0}^N$ be an increasing hierarchy of theories such that $I\Sigma_1 \subseteq T_0$, and for any n, $T_{n+1} \vdash \mathrm{Rfn}(T_n)$. Therefore, there are arithmetical sentences A_1, A_2, ..., A_m such that:*

(i) For any i and j, if $i \neq j$ then $I\Sigma_1 \vdash A_i \wedge A_j \to \bot$

(ii) $I\Sigma_1 \vdash \bigvee_{i=1}^m A_i$

(iii) For any $n \leq N$, and any $i \leq m$, $T_{n+1} \vdash \neg\mathrm{Pr}_{T_n}(\neg A_i)$

(iv) *If we also assume that all theories in the hierarchy are consistent, then for any $n \leq N$ and any $i \leq m$, $\mathbb{N} \vDash \neg \mathrm{Pr}_{T_n}(\neg A_i)$ and $\mathbb{N} \vDash A_m$.*

Proof. First of all, we want to prove the following claim:

Claim. For any increasing reflexive hierarchy $\{T_n\}_{n=0}^{N}$ and any natural number $p \geq 1$, there is another increasing hierarchy $\{T'_n\}_{n=0}^{Np}$ such that for any $n \leq N$, $T'_{np} = T_n$ and for any $i \leq Np-1$, $T'_{i+1} \vdash \mathrm{Cons}(T'_i)$. Moreover, if all of the theories in the T hierarchy are consistent, all of the theories in the T' hierarchy will be consistent, as well.

To prove the claim, define T'_i as follows: For $i = np$, define $T'_i = T_n$, then for the any $np \leq i < (n+1)p-1$ define T'_{i+1} inductively as the theory $T'_i + \mathrm{Cons}(T'_i)$. First of all, we want to show that for any $np \leq i < (n+1)p-1$, $T'_{i+1} \subseteq T'_{(n+1)p}$ and also $T'_{(n+1)p}$ proves the reflection principle for T'_{i+1}. The proof is based on the induction on i. If $i = np$, we know that $T'_{(n+1)p}$ proves the consistency for T'_{np}, hence $T'_{np+1} \subseteq T'_{(n+1)p}$. Moreover, since $T'_{(n+1)p} \vdash \mathrm{Cons}(T'_{np})$, it is easy to check that $T'_{(n+1)p}$ can prove the reflection principle for $T'_{np+1} = T'_{np} + \mathrm{Cons}(T'_{np})$. Suppose that we have the claim for i, and we want to prove it for $i+1$. By IH, $T'_{(n+1)p}$ proves the reflection principle for T'_i, hence it proves the consistency of T'_i and hence $T'_{i+1} \subseteq T'_{(n+1)p}$. Again, it is easy to show that since $T'_{(n+1)p} \vdash \mathrm{Cons}(T'_i)$, $T'_{(n+1)p}$ also proves the reflection principle for $T'_{i+1} = T'_i + \mathrm{Cons}(T'_i)$.

We claim that for any i, $T'_i \subseteq T'_{i+1}$ and T'_{i+1} proves the consistency of T'_i. The proof is based on two different cases of the definition of T'_{i+1}. If we are in the first case, then $i+1 = (n+1)p$ for some n. Then by what we proved so far, the claim is obvious. If we are in the second case, then $T'_{i+1} = T'_i + \mathrm{Cons}(T'_i)$, and hence the claim is again obvious from the definition.
Moreover, if the first hierarchy is consistent, then since all T'_i's are subtheories of $T'_{Np} = T_N$, the second hierarchy is consistent, as well.

It is time to prove the lemma. If $m = 1$, pick $A_1 = (0 = 0)$; then it is easy to verify that this sentence satisfies the conditions of the lemma. The reason is that T_{n+1} proves the consistency of T_n and hence $T_{n+1} \vdash \neg \mathrm{Pr}_n(0 \neq 0)$. Moreover, if all theories are consistent, then $\neg A_1$ is not provable in T_n.
Assume that $m > 1$ and use the hierarchy T from the assumption of the lemma, and also use the aforementioned construction to construct the hierarchy T', for $p = 2m$. We want to define the sentences A_i based on this new hierarchy. Define

$$B_r = \bigvee_{k=1}^{N} (\mathrm{Cons}(T'_{2km-2r}) \wedge \neg \mathrm{Cons}(T'_{2km-2r+1}))$$

for $1 \leq r \leq m-1$. Define $A_1 = B_1$ and $A_r = \bigwedge_{i=1}^{r-1} \neg B_i \wedge B_r$ for $2 \leq r \leq m-1$ and $A_m = \bigwedge_{i=1}^{m-1} \neg B_i$. We claim that these A_i's have the properties in the lemma. First of all, because of the form of A_i's, it is obvious that any two different A_i and A_j are contradictory and also $\bigvee_{r=1}^{m} A_r$. In fact, these claims are first order tautologies and hence they are provable in $I\Sigma_1$. We want to show that

$$T'_{2(n+1)m} \vdash \neg \mathrm{Pr}_{T'_{2nm}}(\neg A_r)$$

We will prove the cases $r \neq 1, m$, $r = 1$ and $r = m$ separately. Assume $r \neq 1, m$. Let us argue in $I\Sigma_1$. If $\neg A_r$ is provable in T'_{2nm}, then by definition $\bigvee_{i=1}^{r-1} B_i \vee \neg B_r$ is provable in T'_{2nm}. From B_t, $t \leq r-1$, we could conclude

$$\bigvee_I (\mathrm{Cons}(T'_{2km-2t})) \vee \bigvee_J (\neg \mathrm{Cons}(T'_{2km-2t+1}))$$

where $I = \{k \mid 2km - 2t + 1 \geq 2nm + 1\}$ and $J = \{k \mid 2km - 2t + 1 < 2nm\}$. First of all, we know that T'_{2nm} proves $\mathrm{Cons}(T'_{2km-2t+1})$ if $k \in J$. The reason is that if $k \in J$, then $2km - 2t + 1 < 2nm$ and since the consistency of any theory is provable in the higher theory in T' hierarchy, we can prove the consistency of $T'_{2km-2t+1}$ in T'_{2nm}. Therefore, we can conclude that the following is provable in T'_{2nm}.

$$\bigvee_I (\mathrm{Cons}(T'_{2km-2t})).$$

On the other hand, we know that if $k \in I$, then $k \geq n+1$ because $2km-2t+1 \geq 2nm + 1$ is impossible when $k \leq n$. Therefore, $2km - 2t \geq 2(n + 1)m - 2t$. Moreover, $2(n + 1)m - 2t \geq 2(n + 1)m - 2(r - 1)$ since $t \leq r - 1$, and since the hierarchy is increasing, $\mathrm{Cons}(T'_{2km-2t})$ implies $\mathrm{Cons}(T'_{2(n+1)m-2(r-1)})$. Hence, B_t implies $\mathrm{Cons}(T'_{2(n+1)m-2(r-1)})$. Furthermore, from

$$\neg B_r = \bigwedge_{k=1}^{N} (\mathrm{Cons}(T'_{2km-2r}) \to \mathrm{Cons}(T'_{2km-2r+1}))$$

we conclude

$$\mathrm{Cons}(T'_{2(n+1)m-2r}) \to \mathrm{Cons}(T'_{2(n+1)m-2r+1}).$$

Therefore, we have

$$T'_{2nm} \vdash (\mathrm{Cons}(T'_{2(n+1)m-2r}) \to \mathrm{Cons}(T'_{2(n+1)m-2r+1})) \vee \mathrm{Cons}(T'_{2(n+1)m-2(r-1)}).$$

Hence

$$T'_{2nm} + \mathrm{Cons}(T'_{2(n+1)m-2r}) \vdash \mathrm{Cons}(T'_{2(n+1)m-2r+1}) \vee \mathrm{Cons}(T'_{2(n+1)m-2(r-1)}).$$

But we have $2(n + 1)m - 2r + 1 \leq 2(n + 1)m - 2(r - 1)$; therefore

$$T'_{2nm} \vdash \mathrm{Cons}(T'_{2(n+1)m-2(r-1)}) \to \mathrm{Cons}(T'_{2(n+1)m-2r+1}).$$

And hence

$$T'_{2nm} + \mathrm{Cons}(T'_{2(n+1)m-2r}) \vdash \mathrm{Cons}(T'_{2(n+1)m-2r+1}).$$

Since $r \leq m$, we have $2(n + 1)m - 2r + 1 \geq 2nm$, therefore we have

$$T'_{2(n+1)m-2r+1} \vdash \mathrm{Cons}(T'_{2(n+1)m-2r+1}).$$

Note that all the parts of this argument is formalizable in $I\Sigma_1$. For the first time we want to use $T'_{2(n+1)m}$ to reach the contradiction. Since $1 \leq r$, then $2(n+1)m-2r+1 < 2(n+1)m$, hence the consistency of $T'_{2(n+1)m-2r+1}$ is provable in $T'_{2(n+1)m}$. Therefore, since we are arguing in $T'_{2(n+1)m}$, we have the consistency of $T'_{2(n+1)m-2r+1}$. On the other hand, we showed

$$\mathrm{Pr}_{T'_{2(n+1)m-2r+1}}(\mathrm{Cons}(T'_{2(n+1)m-2r+1})).$$

By the formalized version of the second incompleteness theorem in $I\Sigma_1$, we know that if a theory proves its own consistency it is inconsistent; hence $T'_{2(n+1)m-2r+1}$ is inconsistent. A contradiction. Therefore, $T'_{2(n+1)m}$ shows that $\neg A_r$ is not provable in T'_{2nm}.

Note that the proof uses the form of $\neg A_r$ which has some positive B_t's and one negative B_r. But Now if we are in the cases $r = 1$ or $r = m$, then $\neg A_r$ has just positive B_t's or just negative B_t's. In these cases it is enough to use the part of the proof which investigates the corresponding B_t's. Again argue in $I\Sigma_1$. For the case, $r = 1$, if T'_{2nm} proves $\neg A_1$, then T'_{2nm} proves $\neg B_1$. Therefore,

$$T'_{2nm} \vdash \bigwedge_{k=1}^{N} (\mathrm{Cons}(T'_{2km-2}) \to \mathrm{Cons}(T'_{2km-1})).$$

Hence

$$T'_{2nm} \vdash (\mathrm{Cons}(T'_{2(n+1)m-2}) \to \mathrm{Cons}(T'_{2(n+1)m-1})).$$

Since $m \geq 1$, we have $2(n+1)m - 1 \geq 2nm$ and hence

$$T'_{2(n+1)m-1} \vdash (\mathrm{Cons}(T'_{2(n+1)m-2}) \to \mathrm{Cons}(T'_{2(n+1)m-1}))$$

and then since $2(n+1)m - 1 > 2(n+1)m - 2$, we have

$$T'_{2(n+1)m-1} \vdash \mathrm{Cons}(T'_{2(n+1)m-1}).$$

Argue in $T'_{2(n+1)m}$. We have the consistency of $T'_{2(n+1)m-1}$. On the other hand, $T'_{2(n+1)m-1}$ proves its own consistency, hence by the formalized second incompleteness theorem, it should be inconsistent. A Contradiction. Therefore, $T'_{2(n+1)m}$ proves that $\neg A_1$ is not provable in T'_{2nm}.

For the proof of the case $r = m$, use the idea of I and J for positive B_t's. It is enough to use I and J, to show that if $\neg A_m$ is provable in T'_{2nm}, then $\mathrm{Cons}(T'_{2(n+1)m-2(m-1)})$ will be provable in $T'_{2(n+1)m-2(m-1)}$. After that, reaching a contradiction is the same as for the other cases.

Since $T'_{2nm} = T_n$, we have a proof for the part (iii). For (iv), if the hierarchy T is consistent, then the hierarchy T' is also consistent and hence if $\neg A_r$ is provable in T'_{2nm} then we have

$$T'_{2(n+1)m-2r+1} \vdash \mathrm{Cons}(T'_{2(n+1)m-2r+1})$$

for cases $1 < r < m$, and

$$T'_{2(n+1)m-1} \vdash \mathrm{Cons}(T'_{2(n+1)m-1})$$

for $r = 1$, and

$$T'_{2(n+1)m-2(m-1)} \vdash \mathrm{Cons}(T'_{2(n+1)m-2(m-1)})$$

for $r = m$. Consider that the arguments for these statements are formalizable in $I\Sigma_1$ and hence they are true. For $1 < r < m$, by the second incompleteness theorem, $T'_{2(n+1)m-2r+1}$ should be inconsistent. A contradiction. Therefore, T'_{2nm} can not prove $\neg A_r$ and hence $T_n \nvdash \neg A_r$. The cases $r = 1, m$ are similar. For the

second part of (iv), note that we know $A_m = \bigwedge_{r=1}^{m} \neg B_r$. We want to show that all B_r's are false. We have

$$B_r = \bigvee_{k=1}^{N} (\mathrm{Cons}(T'_{2km-2r}) \wedge \neg \mathrm{Cons}(T'_{2km-2r+1}))$$

and since the whole T' hierarchy is consistent, all statements $(\mathrm{Cons}(T'_{2km-2r}) \wedge \neg \mathrm{Cons}(T'_{2km-2r+1}))$ are false and hence B_r is false. Then $\neg B_r$ is true and hence A_m is true. □

Lemma 1.6.4. *Let (K, R) be a finite reflexive transitive tree with clusters and let k be one of the nodes in the root cluster. Moreover, let $(\mathbb{N}, \{T_n\}_{n=0}^{N})$ be a reflexive provability model. Then there exists a set of arithmetical sentences $\{S_i\}_{i \in K}$ such that*

(i) If $i \neq j$, $T_0 \vdash S_i \rightarrow \neg S_j$.

(ii) $T_{n+1} \vdash S_i \rightarrow \mathrm{Pr}_n(\bigvee_{(i,j) \in R} S_j)$.

(iii) If $(i, j) \in R$ then $T_{n+1} \vdash S_i \rightarrow \neg \mathrm{Pr}_n(\neg S_j)$.

(iv) $\mathbb{N} \vDash S_k$.

Proof. Define a primitive recursive function $h : \mathbb{N} \rightarrow K$ similar to the h function in the Solovay's proof of the completeness of **GL**.

$$h(0) = k \text{ and } h(x+1) = \begin{cases} j & \text{if } (i,j) \in R \quad \text{and} \quad \mathrm{Prf}_N(x, \neg S_j) \\ h(x) & \text{otherwise} \end{cases}$$

where $S_j = P_{I(j)} \wedge A_j \wedge j = j$ and $P_{I(j)} = \exists y \forall x \geq y \; h(x) \in I(j)$ in which $I(j)$ means the cluster of j. Moreover, A_j's are the sentences constructed in Lemma 1.6.3 for $m = Card(I(j))$ and the hierarchy $\{T_n + P_{I(j)}\}_{n=0}^{N}$. In addition, we choose A_k as the sentence A_m from Lemma 1.6.3. By these sentences, we mean the sentences from the proof of Lemma 1.6.3, and not what the lemma claims. The reason is that we have to be sure that these sentences are definable from the code of the function h which has not been defined yet. The reason is the following:

The function h is going to be defined based on the classical circular argument based on the fixed point lemma in $I\Sigma_1$. The important part is that the A_j's constructed in Lemma 1.6.3 are arithmetical formulas based on the code of $P_{I(j)}$, which makes the whole circular argument possible. It is provably in $I\Sigma_1$ that h is a function. (Note that we put $j = j$ in the definition of S_j to make sure that there is at most one j such that x would be a proof for $\neg S_j$ and this makes the definition of h unambiguous.) It is also provable that h eventually stops in some cluster and since h is a function, this cluster is unique. The existence of such cluster is an obvious application of the fact that h is an increasing function and the tree is finite. Note that all of these facts are provable in $I\Sigma_1$. To prove (i), consider two cases. If i and j belong to different clusters, then $P_{I(i)}$ and $P_{I(j)}$ are contradictory based on what we claimed about the uniqueness of the limit cluster. This contradiction is also provable in $I\Sigma_1$ and hence in T_0. If i and j belong to the same cluster, then by Lemma 1.6.3, we know that A_i and A_j are

contradictory, provable in $I\Sigma_1$, and hence we reach a contradiction for $S_i \wedge S_j$ in T_0. For (ii), we argue in T_{n+1}. If we have S_i, then we have $P_{I(i)}$ and there exists x such that $h(x) \in I(i)$. Since this formula is Σ_1, by Σ_1-completeness we have $\mathrm{Pr}_n(h(x) \in I(i))$. Moreover, h is provably increasing in $I\Sigma_1$ and hence in T_n, and also provably in $I\Sigma_1$ we know that h eventually stops in some cluster, i.e. $\mathrm{Pr}_n(\bigvee_J P_J)$. But we have $\mathrm{Pr}_n(h(x) \in I(i))$. Therefore, the limit should be above i which means $\mathrm{Pr}_n(\bigvee_{(i,j)\in R} P_{I(j)})$. On the other hand, by Lemma 1.6.3 we know that $I\Sigma_1 \vdash \bigvee_{i\in I} A_i$, and we can conclude that $\mathrm{Pr}_n(\bigvee_{(i,j)\in R} P_{I(j)} \wedge A_j)$, hence $\bigvee_{(i,j)\in R} S_j$.

For (iii), we will argue in T_{n+1} and the proof is by contradiction. If we have S_i and $\mathrm{Pr}_n(\neg S_j)$ for some j which $(i,j) \in R$, then there are two possibilities. First, when the clusters of i and j are different. We have $S_i = P_{I(i)} \wedge A_i$, hence we have $P_{I(i)}$ which means that there is some number z, such that for all $y \geq z$, $h(y) \in I(i)$. Moreover, we know that $\mathrm{Pr}_n(\neg S_j)$ and since $T_n \subseteq T_N$, we have $\mathrm{Pr}_N(\neg S_j)$. Therefore, there exists some x such that $\mathrm{Prf}_N(x, \neg S_j)$. It is easy to see that we can pick $x \geq z$. Hence, we can conclude that $h(x+1) \in I(i)$. Since $(i,j) \in R$, j is above all nodes in $I(i)$ and $\mathrm{Prf}_N(x, \neg S_j)$, hence $h(x+1) = j$. But $h(x+1)$ should belong to $I(i)$ and $j \notin I(i)$; a contradiction. Therefore, $\neg\mathrm{Pr}_n(\neg S_j)$.
Assume that the cluster of i and j is I. Then the statement $S_i \to \mathrm{Pr}_n(\neg S_j)$ is equivalent to

$$P_I \wedge A_i \to \mathrm{Pr}_n(P_I \to \neg A_j).$$

Since $\{T_n\}_{n=0}^N$ is a reflexive hierarchy, the hierarchy $\{T_n + P_I\}_{n=0}^N$ is also reflexive. Moreover, A_t's are constructed for this hierarchy, hence by Lemma 1.6.3, we know that

$$T_{n+1} + P_I \vdash \neg\mathrm{Pr}_{T_n+P_I}(\neg A_j)$$

which proves what we wanted.

For (iv), since h eventually stops in some cluster, there is a cluster I, such that $\mathbb{N} \vDash P_I$. If $I \neq I(k)$, since $h(0) = k$, there should be some first element x, such that $h(x) \in I$. Assume $h(x) = i$. Since $x \neq 0$, and $h(x) \neq h(x-1)$, we have $\mathrm{Prf}_N(x-1, \neg S_i)$ and hence, $\mathrm{Pr}_N(P_I \to \neg A_i)$. By Lemma 1.6.3, the theory $T_N + P_I$ should be inconsistent, and therefore we have $T_N \vdash \neg P_I$. On the other hand, the theory T_N is sound, hence $\mathbb{N} \vDash \neg P_I$ which contradicts our assumption. Hence, $I = I(k)$ and therefore, $\mathbb{N} \vDash P_{I(k)}$. On the other hand, $T_N + P_{I(k)}$ is consistent because it is sound, and consequently by Lemma 1.6.3, A_k which was chosen to be the A_m from the lemma, is true; hence $S_k = P_{I(k)} \wedge A_k$ is true. □

The following lemma uses the previous lemma to transfer the truth from a Kripke model to a reflexive provability model.

Lemma 1.6.5. *Assume the conditions of Lemma 1.6.4 and let $\{S_i\}_{i\in K}$ be defined as in that lemma. Define σ as the arithmetical substitution which sends the atom p to $\bigvee_{i\Vdash p} S_i$. For any $i \in K$, any modal formula A and any witness w for A with elements less than N, we have:*

$$\begin{cases} T_{max(w)+1} \vdash S_i \to A^\sigma(w) & \text{if } i \vDash A \\ T_{max(w)+1} \vdash S_i \to \neg A^\sigma(w) & \text{if } i \nvDash A \end{cases}$$

Proof. We prove the lemma by induction on A. If A is an atom and $i \vDash A$, then by the definition we have $T_0 \vdash S_i \to A^\sigma$. If $i \nvDash A$ then all j's in $A^\sigma = \bigvee_{j \vDash A} S_j$ are different from i, and by (i) in Lemma 1.6.4, we conclude $T_0 \vdash S_i \to \neg A^\sigma$. The proof for the boolean cases is easy. For the modal case, if $i \vDash \Box B$, then for all j which $(i,j) \in R$, we have $j \vDash B$. Since w is a witness for $\Box B$, it is equal to (n,u) where n is greater than all the numbers in u. Therefore by IH, $T_{max(u)+1} \vdash S_j \to B^\sigma(u)$ for all j above i. Hence,

$$T_{max(u)+1} \vdash \bigvee_{(i,j)\in R} S_j \to B^\sigma(u).$$

Since $n \geq max(u)+1$, we have

$$T_n \vdash \bigvee_{(i,j)\in R} S_j \to B^\sigma(u).$$

Then

$$I\Sigma_1 \vdash \mathrm{Pr}_n(\bigvee_{(i,j)\in R} S_j \to B^\sigma(u)),$$

and consequently,

$$I\Sigma_1 \vdash \mathrm{Pr}_n(\bigvee_{(i,j)\in R} S_j) \to \mathrm{Pr}_n(B^\sigma(u)).$$

By (ii) in Lemma 1.6.4, we have

$$T_{n+1} \vdash S_i \to \mathrm{Pr}_n(B^\sigma(u)),$$

and $n = max(w)$. Thus, the proof for this case is finished.

If $i \nvDash \Box B$, then there exists j which $(i,j) \in R$ and $j \nvDash B$. Again we have $w = (n,u)$, such that n is greater than all the numbers in u. By IH, $T_{max(u)+1} \vdash S_j \to \neg B^\sigma(u)$. Since $n \geq max(u)+1$,

$$T_n \vdash S_j \to \neg B^\sigma(u)$$

and

$$I\Sigma_1 \vdash \mathrm{Pr}_n(B^\sigma(u) \to \neg S_j)$$

and then

$$I\Sigma_1 \vdash \neg\mathrm{Pr}_n(\neg S_j) \to \neg\mathrm{Pr}_n(B^\sigma(u))$$

and by (iii) in Lemma 1.6.4, we have

$$T_{n+1} \vdash S_i \to \neg\mathrm{Pr}_n(B^\sigma(u))$$

and again since $n = max(w)$, the proof is complete. □

We state and prove the completeness theorem.

Theorem 1.6.6. *(Completeness) Let* $(\mathbb{N}, \{T_n\}_{n=0}^\infty)$ *be a reflexive provability model such that* $(\mathbb{N}, \{T_n\}_{n=0}^\infty) \vDash A$*, then* $\mathbf{S4} \vdash A$*. Therefore, if* $\mathbf{Ref} \vDash A$*, we have* $\mathbf{S4} \vdash A$*.*

Proof. Since $(\mathbb{N}, \{T_n\}_{n=0}^{\infty}) \vDash A$, there are expansions $B_1, \ldots, B_k$ of A and witnesses $w_1, \ldots, w_k$ such that for all arithmetical substitutions σ, we have $\mathbb{N} \vDash \bigvee_{i=0}^{k} B_i^{\sigma}(w_i)$. Define $C = \bigvee_{i=0}^{k} B_i$ and $w = (w_i)_{i=0}^{k}$. Therefore, we know that w is a witness for C in $(\mathbb{N}, \{T_n\}_{n=0}^{\infty})$. We claim that $\mathbf{S4} \vdash C$. Pick N greater than all the numbers in w. If $\mathbf{S4} \nvdash C$ then there exists a finite reflexive transitive tree with clusters (K, R, V), such that in one of the nodes in the root cluster (say k), C is false. Then by Lemmas 1.6.4 and 1.6.5, we can construct an arithmetical substitution, such that $T_{max(w)+1} \vdash S_k \to \neg C^{\sigma}(w)$. Since the model is a reflexive provability model, all T_m's are sound and hence $\mathbb{N} \vDash S_k \to \neg C^{\sigma}(w)$. But by Lemma 1.6.4 we know that $\mathbb{N} \vDash S_k$, thus $\mathbb{N} \vDash \neg C^{\sigma}(w)$, which contradicts with the assumption $\mathbb{N} \vDash C^{\sigma}(w)$. Therefore, $\mathbf{S4} \vdash C$. And finally, since in the presence of the axiom $\mathbf{K}$, all the expansions of a formula are equivalent to the formula itself, we have $\mathbf{S4} \vdash A$.
For the second part of the theorem, it is easy to verify that if $\mathbf{Ref} \vDash A$, then at least for one of the provability models $(\mathbb{N}, \{T_n\}_{n=0}^{\infty})$ we have $(\mathbb{N}, \{T_n\}_{n=0}^{\infty}) \vDash A$. And then the claim follows from the first part. □

1.6.3 Uniform and Strong Completeness

In this subsection we will strengthen the completeness theorem of the last subsection to a stronger version of uniform strong completeness theorem. The proof will be just the uniform version of the previous completeness proof. Therefore, first of all we need a uniform version of Lemma 1.6.3.

Definition 1.6.7. *A hierarchy $\{T_n\}_{n=0}^{\infty}$ of theories is called uniform if there exists a Σ_1 formula $\mathrm{Prf}(x, y, z)$ such that for any n, m and A, $\mathrm{Prf}(n, m, \ulcorner A \urcorner)$ iff m is a code of a proof for A in T_n. The hierarchy is called uniformly increasing if it is a uniform hierarchy and also we have $I\Sigma_1 \subseteq T_0$ provably in $I\Sigma_1$ and $I\Sigma_1 \vdash \forall x \forall z (\exists y\ \mathrm{Prf}(x, y, z) \to \exists w\ \mathrm{Prf}(x+1, w, z))$. And finally it is called uniformly reflexive hierarchy if it is a uniformly increasing hierarchy such that for any formula A, $I\Sigma_1 \vdash \forall x \exists y\, \mathrm{Prf}(x+1, y, \exists w\, \mathrm{Prf}(x, w, A) \to A)$.*

Lemma 1.6.8. *Let $\{T_n\}_{n=0}^{\infty}$ be a uniformly reflexive hierarchy of theories. Then, there is an arithmetical sentence $A(x, y)$ such that:*

(i) $I\Sigma_1 \vdash \forall x, z \leq y\ (x \neq z \wedge A(x, y) \wedge A(z, y) \to \bot)$

(ii) *For all m,* $I\Sigma_1 \vdash \bigvee_{i=1}^{m} A(i, m)$

(iii) *For any n, and any $i \leq m$,* $T_{n+1} \vdash \neg \mathrm{Pr}_{T_n}(\neg A(i, m))$

(iv) *If we also assume that all theories in the hierarchy are consistent, then for any n, and any $i \leq m$,* $\mathbb{N} \vDash \neg \mathrm{Pr}_{T_n}(\neg A(i, m))$ *and* $\mathbb{N} \vDash A(m, m)$.

Proof. The proof is basically the same as the proof of Lemma 1.6.3. The only difference is that, here we have to define everything uniformly. First of all we need to define the hierarchy T'. Since T is a uniformly reflexive hierarchy, it is easy to prove that the hierarchy T' is a uniform hierarchy. Note that the definition of this new hierarchy is also uniform in p, i.e. there exists a proof predicate $\mathrm{Prf}(x, y, z, t)$

which means that y is a proof for z in T'_x when we choose t as our p. Define, $B(x,y)$ as the following:

$$B(x,y) = \exists z \geq 1\ (\mathrm{Cons}(T'_{2zy-2x}) \wedge \neg \mathrm{Cons}(T'_{2zy-2x+1})),$$

and

$$A(x,y) = \forall 1 \leq z \leq x-1\ \neg B(z,y) \wedge B(x,y).$$

Note that $A(x,y)$ and $B(x,y)$ are the uniform versions of A_r and B_r in which x stands for the index r and y for the number m. The proof of the properties we claimed is exactly same as the proof of Lemma 1.6.3. The reason is that all properties are based on the standard numbers n, i and m. The only exception is (i), which is easily proved from the definition. □

Theorem 1.6.9. *(Uniform Completeness) Let $\{T_n\}_{n=0}^{\infty}$ be a uniform reflexive hierarchy of sound theories. Then there exists an arithmetical substitution $*$, such that for any modal formula A, if there exists a witness w such that for all $M \vDash \bigcup_n T_n$, $(M, \{T_n\}_{n=0}^{\infty}) \vDash A^*(w)$ then $\mathbf{S4} \vdash A$.*

Proof. First, note that according to the filtration method (see [10]), there exists a primitive recursive algorithm which reads A as an input and constructs a counter model (finite transitive reflexive tree with clusters) for A if $\mathbf{S4} \nvdash A$, and outputs zero, otherwise. Call this primitive recursive function, f. Therefore, if we use A_a to emphasize that the code for A is a, we have $f(a) = (W_a, R_a, V_a, w_a)$ in which w_a is a node in the root cluster such that $w_a \nvDash A_a$. The reason why such an f exists is that the size of a counter model is elementary bounded by the size of the code of the formula. (See [10].) Assume that the function $\langle \cdot, \cdot \rangle$ is some canonical pairing function which is primitive recursive. Define $g(a)$ as the following primitive recursive function: Compute $f(a)$, change the name of all nodes w in W_a to $\langle w, a \rangle$ and code the whole model again.
Pick all $g(a)$'s and put all of them over one new reflexive root, k; and for valuation, use the induced valuation of the model plus the fact that the node k does not accept any atom. Then, use the technique of Lemma 1.6.4 and define the function h on the whole new model:

$$h(0) = k \text{ and } h(x+1) = \begin{cases} j & \text{if } R(h(x), z) \quad \text{and} \quad \mathrm{Prf}_T(x, \neg S(z)) \\ h(x) & \text{otherwise} \end{cases}$$

Where firstly, $T = \bigcup_{n=0}^{\infty} T_n$. It is easy to check that since the hierarchy is uniform, its union is also a recursively enumerable theory which has the following property: $I\Sigma_1 \vdash \mathrm{Pr}_n(A) \to \mathrm{Pr}_T(A)$. Secondly, $R(y,z)$ is a primitive recursive relation (Δ_1 formula in $I\Sigma_1$) which reads nodes y and z and if $y \neq k$, it decides whether they belong to the same model $g(pr_0(z))$, and if yes, whether (y,z) belongs to the relation of $g(pr_0(z))$, i.e. $R_{g(pr_0(z))}$. And if $y = k$, then the relation $R(y,z)$ decides whether z is in the $g(pr_0(z))$ or not (where $pr_0(z)$ is the index of the model which z belongs to). This R is a formalization of the accessibility relation of the new model. Note that we have to choose R in a way that the following holds:

(i) $I\Sigma_1 \vdash \forall x, y, z\ (R(x,y) \wedge R(y,z) \to R(x,z))$

(ii) For any node $i \neq k$, $I\Sigma_1 \vdash \forall x (R(i,x) \to \bigvee_{R_{g(pr_0(i))}(i,j)} x = j)$

It is easy to find such an R. The idea is, first using g to define a primitive recursive function $H(z)$ which reads z and outputs the whole set above z. Then define $R(x, y)$ as the existence of a sequence w from x to y such that for any r, w_{r+1} belongs to $H(w_r)$. The proof for these two properties are starightforward. (i) holds because of our transitive definition of R. (ii) needs the claim that if w is a sequence from i to x, then $x \in H(i)$. Use induction on the length of w to prove the claim.

And finally, the formula

$$S(z) = \exists y \forall x \geq y h(x) \in I(z) \wedge A(z, Card(I(z))) \wedge z = z$$

where $I(z)$ is a primitive recursive function, which reads z and computes the whole cluster of z. Note that here we use a uniform version of S_i's, and consequently we need the uniform version of A_r's. For any $i \neq k$, the model above w_i is a finite reflexive transitive tree with clusters, and hence with the same arguments, we have the following:

(i) $T_0 \vdash \forall x, y\ (x \neq y \to (S(x) \to \neg S(y)))$.

(ii) $T_{n+1} \vdash S(i) \to \mathrm{Pr}_n(\bigvee_{(i,j)\in R} S(j))$ for all $i \neq k$.

(iii) If $(i, j) \in R$ then $T_{n+1} \vdash S_i \to \neg \mathrm{Pr}_n(\neg S_j)$ for all i.

(iv) $\mathbb{N} \vDash S_k$.

Since the model above any node $i \neq k$ is a finite model, the proof is the same as the proof of Lemma 1.6.4, with only some minor changes. Firstly, for (i), we need the uniform version of the proof of Lemma 1.6.4. It is implied by the facts that h is a provably total function in $I\Sigma_1$ and also the part (i) in Lemma 1.6.8. Secondly, for (ii), we need to prove that if the function reaches i, then the limit cluster exists and it is above the cluster $I(i)$. It is provable in $I\Sigma_1$. The idea is based on the fact that h is increasing and also the fact that if h reaches i, we can find the elements above i. These simple facts are provable by two properties of R which are mentioned before.

Define the arithmetical substitution as follows: $p^* = \exists z\ S(z) \wedge V(z, p)$ where $V(z, p)$ is a primitive recursive predicate (i.e. a Δ_1 formula in $I\Sigma_1$) which reads z and p and if $z \neq k$ decides whether p is true in the node z in the model $g(a)$, where $a = pr_0(z)$ is the index of the model which z belongs to. And if $z = k$, then rejects for all p. Since g is primitive recursive, this primitive recursive predicate exists. Note that V is a formalization of the valuation of the new model.
By a similar proof of Lemma 1.6.5 we know that for all $i \neq k$, we have

$$\begin{cases} T_{max(w)+1} \vdash S_i \to A^\sigma(w) & \text{if } i \vDash A \\ T_{max(w)+1} \vdash S_i \to \neg A^\sigma(w) & \text{if } i \nvDash A \end{cases}$$

If $\mathbf{S4} \nvdash A$, then $i = w_a \nvDash A$, where a is the code of A. We have

$$T_{max(w)+1} \vdash S_i \to \neg A^*(w).$$

Hence for all $n \geq max(w) + 1$,

$$T_n \vdash S_i \to \neg A^*(w).$$

Then by

$$T_{n+1} \vdash S_k \to \neg \mathrm{Pr}_n(\neg S_i),$$

we have

$$T_{n+1} \vdash S_k \to \neg \mathrm{Pr}_n(A^*(w)).$$

Since T_{n+1} is sound, $\mathbb{N} \vDash \neg \mathrm{Pr}_n(A^*(w))$ which means $T_n \nvdash A^*(w)$, and since n could be any sufficiently large number, $T \nvdash A^*(w)$, therefore, there is M, a model of $T = \bigcup_n T_n$, such that $M \nvDash A^*(w)$, which is a contradiction. Hence, $\mathbf{S4} \vdash A$. □

Using the previous lemma, we are able to prove the strong completeness theorem.

Theorem 1.6.10. *(Uniform Strong Completeness) Let $\{T_n\}_{n=0}^{\infty}$ be a uniformly reflexive hierarchy of sound theories. Then there exists an arithmetical substitution $*$, such that for any modal sequent $\Gamma \Rightarrow A$, if there exist witnesses u and v such that for all $M \vDash \bigcup_n T_n$, $(M, \{T_n\}_{n=0}^{\infty}) \vDash \Gamma^*(u) \Rightarrow A^*(v)$, then $\mathbf{S4} \vdash \Gamma \Rightarrow A$. Moreover, if $\mathbf{Ref} \vDash \Gamma \Rightarrow A$, then $\Gamma \vdash_{\mathbf{S4}} A$.*

Proof. Use the arithmetical substitution from the uniform completeness. Since

$$(M, \{T_n\}_{n=0}^{\infty}) \vDash \Gamma^*(u) \Rightarrow A^*(v)$$

for all $M \vDash \bigcup_n T_n$, then $\bigcup_n T_n + \Gamma^*(u) \vdash A^*(v)$. Therefore, there is a finite subset $\Delta \subseteq \Gamma$ and a witness w, a subset of u, such that $\bigcup_n T_n + \Delta^*(w) \vdash A^*(v)$. Thus, for all $M \vDash \bigcup_n T_n$, we have

$$(M, \{T_n\}_{n=0}^{\infty}) \vDash \Delta^*(u) \Rightarrow A^*(v).$$

By uniform completeness, we have $\mathbf{S4} \vdash \Delta \Rightarrow A$ and hence, $\mathbf{S4} \vdash \Gamma \Rightarrow A$.
The second part of the theorem, is obvious from the first part; because if $\mathbf{Ref} \vDash \Gamma \Rightarrow A$, then the assumption of the first part is true for some sequence of expansions $\bar{\Gamma}$ and $B_1, B_2, \ldots, B_r$. Hence $\bar{\Gamma} \vdash_{\mathbf{S4}} \bigvee_{i=0}^{r} B_i$. Since in the presence of the axiom $\mathbf{K}$, the expansions of a formula are equivalent to the formula itself, we have $\Gamma \vdash_{\mathbf{S4}} A$. □

1.7 The Logics GL and GLS

As Solovay showed in his pioneering work, [20], the logic **GL** is sound and complete for the interpretation that interprets all boxes as provability predicates in some appropriate theory. Moreover, he showed that if we change the definition slightly, we can also capture the logic **GLS**. We translate his results into our framework and after defining constant and sound-constant provability models, we will show the soundness and completeness of **GL** and **GLS** for the classes of all constant provability models and all sound-constant provability models, respectively. In fact, the soundness-completeness theorems of these logics are just a new representation of Solovay's results. Consequently, we can claim that our provability interpretation is actually a generalization of Solovay's provability interpretation.

1.7.1 The Case GL

First of all the definition of the constant and sound-constant provability models:

Definition 1.7.1. *A provability model, $(M, \{T_n\}_{n=0}^{\infty})$ is constant if for any n and m, $(M, \{T_n\}_{n=0}^{\infty})$ thinks that $T_n = T_m$, i.e. $M \vDash \mathrm{Pr}_{T_m}(A) \leftrightarrow \mathrm{Pr}_{T_n}(A)$ and $M \vDash \mathrm{Pr}_{T_0}(\mathrm{Pr}_{T_m}(A) \leftrightarrow \mathrm{Pr}_{T_n}(A))$ for all sentences A; and it is called a sound-constant model when it is constant and for any n, M thinks that T_n is sound, i.e. $M \vDash \mathrm{Pr}_{T_n}(A) \rightarrow A$ for any sentence A. The class of all constant provability models and the class of all sound-constant provability models will be denoted by* **Cst** *and* **sCst**, *respectively.*

Remark 1.7.2. *In the previous definition we used a notion for the equality of the-ories which seems ad-hoc and artificial. Here in this remark, we will justify that definition. Intuitively, M thinks that two theories are equal, when their provability properties are the same. In a more precise way, we say that M thinks T_n and T_m are equal, when for any modal sentence $\phi(p)$, any witness w and any arithmetical substitution σ for all atoms except p, $M \vDash \phi^{\sigma}(\mathrm{Pr}_m(A))(w) \leftrightarrow \phi^{\sigma}(\mathrm{Pr}_n(A))(w)$. We will show that this definition of equality is equivalent to the original one. First of all, if we use $\phi(p) = p$, we will have $M \vDash \mathrm{Pr}_{T_m}(A) \leftrightarrow \mathrm{Pr}_{T_n}(A)$. More-over, if we use $\phi(p) = \Box(p \leftrightarrow q)$, $w = (0)$ and σ where $q^{\sigma} = \mathrm{Pr}_n(A)$, we have $M \vDash \mathrm{Pr}_{T_0}(\mathrm{Pr}_{T_m}(A) \leftrightarrow \mathrm{Pr}_{T_n}(A))$. For the converse, we use induction on ϕ to show the following claim.*

Claim. *For any formula $\phi(p)$, any witness w and any arithmetical substitution σ for all atoms except p, M thinks that both of the following state-ments are true: $\phi^{\sigma}(\mathrm{Pr}_m(A))(w) \leftrightarrow \phi^{\sigma}(\mathrm{Pr}_n(A))(w)$ and $T_0 \vdash \phi^{\sigma}(\mathrm{Pr}_m(A))(w) \leftrightarrow \phi^{\sigma}(\mathrm{Pr}_n(A))(w)$.*

The atomic case and the boolean case are obvious. For the modal case, it is an easy consequence of the fact that Σ_1-completeness and some basic facts about the provability predicate are true in M.

We are ready to prove the soundness-completeness result for **GL**. First of all, a technical lemma.

Lemma 1.7.3. *Let $(M, \{T_n\}_{n=0}^{\infty})$ be a constant provability model. Then for any modal formula A, any witness w and any arithmetical substitution σ, if $\mathbf{0}$ assigns zero to all the boxes of A, then M thinks that both of the following statements are true: $A^{\sigma}(w) \leftrightarrow A^{\sigma}(\mathbf{0})$ and $T_0 \vdash A^{\sigma}(w) \leftrightarrow A^{\sigma}(\mathbf{0})$.*

Proof. Use induction on A. The case for the atoms and the boolean connectives are easy. For the modal case, if $A = \Box B$, and $w = (n, u)$, then by IH, M thinks $T_0 \vdash B^{\sigma}(u) \leftrightarrow B^{\sigma}(\mathbf{0})$. Hence $T_n \vdash B^{\sigma}(u) \leftrightarrow B^{\sigma}(\mathbf{0})$ and by Σ_1-completeness, $M \vDash \mathrm{Pr}_n(B^{\sigma}(u) \leftrightarrow B^{\sigma}(\mathbf{0}))$. Thus $\mathrm{Pr}_n(B^{\sigma}(u)) \leftrightarrow \mathrm{Pr}_n(B^{\sigma}(\mathbf{0}))$ is true in M. Since $\mathrm{Pr}_n(B^{\sigma}(\mathbf{0}))$ and $\mathrm{Pr}_0(B^{\sigma}(\mathbf{0}))$ are equivalent in M, we have

$$M \vDash \mathrm{Pr}_n(B^{\sigma}(u)) \leftrightarrow \mathrm{Pr}_0(B^{\sigma}(\mathbf{0})).$$

For the other part of the claim, for $\Box B$, we have $M \vDash \mathrm{Pr}_n(B^{\sigma}(u) \leftrightarrow B^{\sigma}(\mathbf{0}))$. Therefore by Σ_1-completeness, M thinks $T_0 \vdash \mathrm{Pr}_n(B^{\sigma}(u) \leftrightarrow B^{\sigma}(\mathbf{0}))$. Hence $T_0 \vdash$

$\Pr_n(B^\sigma(u)) \leftrightarrow \Pr_n(B^\sigma(\mathbf{0}))$ is true in M. But we know that M thinks that

$$T_0 \vdash \Pr_n(B^\sigma(\mathbf{0})) \leftrightarrow \Pr_0(B^\sigma(\mathbf{0})),$$

therefore, M thinks that

$$T_0 \vdash \Pr_n(B^\sigma(u)) \leftrightarrow \Pr_0(B^\sigma(\mathbf{0})).$$

□

Theorem 1.7.4. *(Soundness) If* $\Gamma \vdash_{\mathbf{GL}} A$*, then* $\mathbf{Cst} \vDash \Gamma \Rightarrow A$.

Proof. If $\Gamma \vdash_{\mathbf{GL}} A$ then there exists a finite $\Delta \subseteq \Gamma$ such that $\mathbf{GL} \vdash \bigwedge \Delta \to A$. Then by Theorem 1.2.3, we have $I\Sigma_1 \vdash \Delta^\sigma(\mathbf{0}) \to A^\sigma(\mathbf{0})$. Thus for any model M, $M \vDash \Gamma^\sigma(\mathbf{0}) \Rightarrow A^\sigma(\mathbf{0})$. Pick any arbitrary witnesses for Γ and A say w_Γ and w_A. By using the Lemma 1.7.3 we will have $M \vDash \Gamma^\sigma(w_\Gamma) \Rightarrow A^\sigma(w_A)$. □

For the completeness of **GL** we have:

Theorem 1.7.5. *(Uniform Strong Completeness) Let* $I\Sigma_1 \subseteq T$ *be an r.e.* Σ_1*-sound theory and* $\{T_n\}_{n=0}^\infty$ *be a hierarchy of theories such that for any* n*,* $T_n = T$*, then there is an arithmetical substitution* $*$ *such that for any modal sequent* $\Gamma \Rightarrow A$*, if for all* $M \vDash T$*, we have* $(M, \{T_n\}_{n=0}^\infty) \vDash \Gamma \Rightarrow A$*, then* $\Gamma \vdash_{\mathbf{GL}} A$*. In particular, if* $\mathbf{Cst} \vDash \Gamma \Rightarrow A$ *then* $\Gamma \vdash_{\mathbf{GL}} A$.

Proof. Pick $*$ as the uniform arithmetical substitution in Solovay's completeness theorem for T (see Preliminaries and [7]). Pick $M \vDash T$, arbitrarily. We have $(M, \{T_n\}_{n=0}^\infty) \vDash \Gamma \Rightarrow A$, hence there are a sequence of expansions $\bar{\Gamma}$ and expansions $\{A_i\}_{i=0}^r$ of A and witnesses u and w_i such that

$$M \vDash \bar{\Gamma}^*(u) \Rightarrow \bigvee_{i=0}^r A_i^*(w_i).$$

Since all the theories are equal, we can easily verify that for any formula B and any witness v, $B^*(v)$ is equivalent to B^*, where B^* means a combination of substituting all the atoms by $*$ and interpreting any box as the provability predicate for T. Then we have

$$M \vDash \bar{\Gamma}^* \Rightarrow \bigvee_{i=0}^r A_i^*.$$

Moreover, it is easy to prove that if B is an expansion of C, then B^* is equivalent to C^* in $I\Sigma_1$ and hence $M \vDash \Gamma^* \Rightarrow A^*$. Since M is arbitrary, we have $T + \Gamma^* \vdash A^*$, therefore, there is a finite subsequence $\Delta \subseteq \Gamma$ such that $T + \Delta^* \vdash A^*$. Then by Solovay's uniform completeness theorem, we have $\Delta \vdash_{\mathbf{GL}} A$, thus $\Gamma \vdash_{\mathbf{GL}} A$. For the second part of the theorem, it is easy to show that if $\mathbf{Cst} \vDash \Gamma \Rightarrow A$, then the assumption of the first part for $T = I\Sigma_1$ is met, and hence $\Gamma \vdash_{\mathbf{GL}} A$. □

1.7.2 The Case GLS

For the case of **GLS** we have:

Theorem 1.7.6. *(Soundness) If* $\Gamma \vdash_{\mathbf{GLS}} A$*, then* $\mathbf{sCst} \vDash \Gamma \Rightarrow A$.

Proof. If $\Gamma \vdash_{\mathbf{GLS}} A$, then there are formulas $B_1, B_2, \ldots, B_k$ such that $\Gamma \vdash_{\mathbf{GL}} \bigwedge_{i=1}^{k}(\Box B_i \to B_i) \to A$. By the proof of the soundness of **GL**, we know that for any constant provability model and any arithmetical substitution σ, $M \vDash \Gamma^\sigma(\mathbf{0}) + \bigwedge_{i=1}^{k}(\mathrm{Pr}_0(B_i^\sigma(\mathbf{0})) \to B_i^\sigma(\mathbf{0})) \Rightarrow A^\sigma(\mathbf{0})$. Since $M \vDash \mathrm{Pr}_0(\phi) \to \phi$ for any arithmetical ϕ, we have $M \vDash \Gamma^\sigma(\mathbf{0}) \Rightarrow A^\sigma(\mathbf{0})$. Use Lemma 1.7.3 to change the index of the theories from zero to any arbitrary witness. □

Moreover, we have the completeness theorem.

Theorem 1.7.7. *(Completeness) Let* $I\Sigma_1 \subseteq T$ *be a sound r.e. theory and* $\{T_n\}_{n=0}^{\infty}$ *be a hierarchy of theories such that for any* n, $T_n = T$*. If* $(\mathbb{N}, \{T_n\}_{n=0}^{\infty}) \vDash A$*, then* $\mathbf{GLS} \vdash A$*; and especially, if* $\mathbf{sCst} \vDash A$*, then* $\mathbf{GLS} \vdash A$.

Proof. By the assumption, we have $(\mathbb{N}, \{T_n\}_{n=0}^{\infty}) \vDash A$. Hence, there are expansions $\{A_i\}_{i=0}^{r}$ of A and witnesses w_i such that for all arithmetical substitutions σ, $\mathbb{N} \vDash \bigvee_{i=0}^{r} A_i^\sigma(w_i)$. Since all the theories are equivalent, it is easy to show that for any formula B and any witness v, $B^\sigma(v)$ is equivalent to B^σ, where B^σ means a combination of substituting any atom by σ and interpreting any box as the provability predicate for T. Therefore, $\mathbb{N} \vDash \bigvee_{i=0}^{r} A_i^\sigma$. Moreover, it is easy to prove that if B is an expansion of C, then B^σ is equivalent to C^σ in $I\Sigma_1$, hence $\mathbb{N} \vDash A^\sigma$. Since σ is arbitrary, based on Solovay's second completeness theorem, $\mathbf{GLS} \vdash A$. For the second part of the theorem, it is easy to verify that if $\mathbf{sCst} \vDash A$ then the assumption of the first part for $T = I\Sigma_1$ is met and hence $\mathbf{GLS} \vdash A$. □

1.8 The Extensions of KD45

Intuitively, the logic **S5** does not admit any provability interpretation. The informal reason is as follows: The axiom $\mathbf{5} : \neg\Box A \to \Box\neg\Box A$ simply states that if A is not provable in a theory T_n, then this fact will be provable in T_{n+1}, i.e.

$$T_n \nvdash A \Rightarrow T_{n+1} \vdash \neg\mathrm{Pr}_n(A).$$

Moreover, the axiom **T** asserts that all theories are sound, hence

$$T_n \nvdash A \Leftrightarrow T_{n+1} \vdash \neg\mathrm{Pr}_n(A).$$

We can use the last equivalence and the fact that the theory T_{n+1} is recursively enumerable to find a decision procedure for the provability in the theory $I\Sigma_1 \subseteq T_n$, which is impossible.
The above argument is based on the axiom **5** and the fact that all theories are sound. But it is possible to weaken the soundness part to a certain consistency assumption which generalizes the above argument to all extensions of the logic **KD45**.

Theorem 1.8.1. *There is no provability model* $(M, \{T_n\}_{n=0}^{\infty})$ *such that*

$$(M, \{T_n\}_{n=0}^{\infty}) \vDash \mathbf{KD45}.$$

Hence, there are no provability models for any extension of the logic **KD45**. *In particular,* **S5** *does not have any provability interpretation.*

Proof. The proof we present here is more complex than the natural proof of this theorem, because we use weaker assumptions than what is available in **KD45**. The reason of our interest in this more complex proof is that we will use the same proof for the case of the classical propositional logic, and in that case we just have access to these weaker assumptions.
We prove the claim by contradiction. Suppose that there is a provability model $(M, \{T_n\}_{n=0}^{\infty})$ such that $(M, \{T_n\}_{n=0}^{\infty}) \vDash \mathbf{KD45}$. First, we show that the following three statements are true in M, then we will use these statements to reach the contradiction.

(*i*) For any n, M thinks that $T_{n+1} \nvdash \mathrm{Pr}_n(\bot)$. (Weak version of the consistency assumption.)

(*ii*) For any n, there exist $N > n$ and $s < N$ such that M thinks that $T_N \vdash \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot)) \to \mathrm{Pr}_s(\bot)$. (Weak version of the provability of the consistency assumption.)

(*iii*) There are m, n and k such that M thinks that for any arithmetical statement ϕ,

$$\neg\mathrm{Pr}_n(\phi) \to \mathrm{Pr}_{m+1}(\mathrm{Pr}_k(\phi) \to \mathrm{Pr}_m(\bot)).$$

(Weak version of the axiom **5**).

To prove (*i*), for any number n, define $\Box^n\top$ as follows: $\Box^0\top = \top$ and $\Box^{n+1}\top = \Box\Box^n\top$. Consider the formula $\neg\Box\Box(\bot \wedge \Box^n\top)$, which is a theorem of **KD45**. Therefore, we have expansions of this formula, of the form $\neg\Box\bigvee_{j=0}^{s_i}\Box\bigvee_{k=0}^{t_{ij}}(\bot \wedge B_{ijk})$ for $0 \leq i \leq r$, where B_{ijk} is an expansion of $\Box^n\top$. Moreover, there are witnesses $w_i = (n_i, (m_{ij}, (u_{ijk})_{k=0}^{t_{ij}})_{j=0}^{s_i})$ for any of these expansions such that for any arithmetical substitution σ, we have

$$M \vDash \bigvee_{i=0}^{r} \neg\Box \bigvee_{j=0}^{s_i} \Box(\bigvee_{k=0}^{t_{ij}} (\bot \wedge B_{ijk}))^{\sigma}(w_i).$$

Since the number of the boxes in $\Box^n\top$ is n, and witnesses for these boxes should be increasing, we have $m_{ij} \geq n$ and hence $n_i \geq n+1$. Define $M = min_{ij}(m_{ij})$ and $N = min_i(n_i)$. Since B_{ijk} is an expansion of the theorem $\Box^n\top$, we can easily show that $B_{ijk}(u_{ijk})$ is provable in $I\Sigma_1$. Hence, it is easy to see that $M \vDash \neg\mathrm{Pr}_N(\mathrm{Pr}_M(\bot))$ for some $N > M \geq n$. Therefore, if $M \vDash \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$, and since $N > M \geq n$, we have $\mathrm{Pr}_N(\mathrm{Pr}_M(\bot))$, which is a contradiction.

For (*ii*), apply the same method to the formula $\Box(\Box\Box(\Box\bot \wedge \Box^n\top) \to \Box\bot)$ which is again a theorem of **KD45**. Then there are expansions of the form

$\Box\bigvee_{j=0}^{q_j}(\Box(\bigvee_{k=0}^{p_{ij}}\Box\bigvee_{l=0}^{t_{ijk}}B_{ijkl})\to\Box\bot)$ where B_{ijkl} is an expansion of $\Box\bot\wedge\Box^n\top$ and there are witnesses $w_i=(n_i,(m_{ij},(r_{ijk},(u_{ijkl})_{l=0}^{t_{ijk}})_{k=0}^{p_{ij}},s_{ij})_{j=0}^{q_i})$ such that

$$M\vDash\bigvee_{i=0}^{r}(\Box(\bigvee_{j=0}^{q_j}(\Box(\bigvee_{k=0}^{p_{ij}}\Box\bigvee_{l=0}^{t_{ijk}}B_{ijkl})\to\Box\bot))^{\sigma}(w_i).$$

Once more, with the same reasoning as in the case (i), $n\leq r_{ijk}<m_{ij}<n_i$. Define $N=max_i(n_i)$, $r=min_{ijk}(r_{ijk})$, $m=min_{ij}(m_{ij})$ and $s=max_i(s_i)$. Hence $N>m,r,s$ and $m>r\geq n$. Since the theories in the hierarchy $\{T_n\}_{n=0}^{\infty}$ are provably increasing, it is easy to prove

$$M\vDash\mathrm{Pr}_N(\mathrm{Pr}_m(\mathrm{Pr}_r(\bot))\to\mathrm{Pr}_s(\bot)).$$

Because $m>r\geq n$, we have

$$M\vDash\mathrm{Pr}_N(\mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))\to\mathrm{Pr}_s(\bot)).$$

Since n is arbitrary, we have proved that for any n, there exists $N>n$, $s<N$ such that

$$M\vDash\mathrm{Pr}_N(\mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))\to\mathrm{Pr}_s(\bot)),$$

and this is what we wanted.

For (iii) we know that $\neg\Box p\to\Box(\Box p\to\Box\bot)$ is provable in **KD45** and consequently it is true in the model. Therefore, there are some expansions of the formula $\neg\Box p\to\Box\bigvee_{j=0}^{s_i}(\Box p\to\Box\bot)$, and some witnesses $(n_i,m_i,(k_{ij},l_{ij})_{j=0}^{s_i})$ for them, such that for any arithmetical substitution σ,

$$M\vDash\bigvee_{i=0}^{r}(\neg\Box p\to\Box\bigvee_{j=0}^{s_i}(\Box p\to\Box\bot))^{\sigma}(n_i,m_i,(k_{ij},l_{ij})_{j=0}^{s_i}).$$

Define $n=max_i(n_i)$, $k=min_{ij}(k_{ij})$, $m=max_i(k_i)$ and $l=max_{ij}(l_{ij})$. It is easy to show that

$$M\vDash\neg\mathrm{Pr}_n(p^{\sigma})\to\mathrm{Pr}_m(\mathrm{Pr}_k(p^{\sigma})\to\mathrm{Pr}_l(\bot)).$$

It is easily verified that we can increase m and l; therefore, w.l.o.g. we can assume that $m=l+1$. Send p to ϕ to prove the claim, and this completes the proof of the statement (iii).

For the proof of Theorem 1.8.1, we want to use these three statements to reach a contradiction. First of all, to simplify the proof, use the following notation. For any a and b, define the theory $T_{b_a}=T_b+\mathrm{Cons}(T_a)$. Thus, by $\mathrm{Pr}_{b_a}(A)$, we mean $\mathrm{Pr}_{T_{b_a}}$. Now, (iii) would be equivalent to

$$M\vDash\neg\mathrm{Pr}_n(p^{\sigma})\to\mathrm{Pr}_{m_l}(\neg\mathrm{Pr}_k(p^{\sigma})).$$

Put $\phi=\mathrm{Pr}_{m_l}(\bot)$; therefore,

$$M\vDash\neg\mathrm{Pr}_n(\mathrm{Pr}_{m_l}(\bot))\to\mathrm{Pr}_{m_l}(\neg\mathrm{Pr}_k(\mathrm{Pr}_{m_l}(\bot))).$$

On other hand by the formalized Σ_1-completeness, we have

$$I\Sigma_1\vdash\neg\mathrm{Pr}_k(\mathrm{Pr}_{m_l}(\bot))\to\neg\mathrm{Pr}_{m_l}(\bot),$$

hence,

$$T_{m_l} \vdash \neg \mathrm{Pr}_k(\mathrm{Pr}_{m_l}(\bot)) \to \neg \mathrm{Pr}_{m_l}(\bot).$$

Moreover, by Σ_1-completeness, we have

$$I\Sigma_1 \vdash \mathrm{Pr}_{m_l}(\neg \mathrm{Pr}_k(\mathrm{Pr}_{m_l}(\bot)) \to \neg \mathrm{Pr}_{m_l}(\bot)).$$

Therefore,

$$I\Sigma_1 \vdash \mathrm{Pr}_{m_l}(\neg \mathrm{Pr}_k(\mathrm{Pr}_{m_l}(\bot))) \to \mathrm{Pr}_{m_l}(\neg \mathrm{Pr}_{m_l}(\bot)).$$

And since $M \vDash I\Sigma_1$, we have

$$M \vDash \neg \mathrm{Pr}_n(\mathrm{Pr}_{m_l}(\bot)) \to \mathrm{Pr}_{m_l}(\neg \mathrm{Pr}_{m_l}(\bot)).$$

Based on Gödel's second incompleteness theorem formalized in $I\Sigma_1$, we can conclude

$$I\Sigma_1 \vdash \neg \mathrm{Pr}_{m_l}(\bot) \to \neg \mathrm{Pr}_{m_l}(\neg \mathrm{Pr}_{m_l}(\bot)).$$

However, by (i), we have

$$M \vDash \neg \mathrm{Pr}_{l+1}(\mathrm{Pr}_l(\bot)),$$

hence $M \vDash \neg \mathrm{Pr}_{m_l}(\bot)$. Since $M \vDash I\Sigma_1$,

$$M \vDash \neg \mathrm{Pr}_{m_l}(\neg \mathrm{Pr}_{m_l}(\bot)).$$

Therefore,

$$M \vDash \mathrm{Pr}_n(\mathrm{Pr}_{m_l}(\bot)),$$

and thus by definition of T_{m_l} we have

$$M \vDash \mathrm{Pr}_n(\mathrm{Pr}_m(\mathrm{Pr}_l(\bot))).$$

By (ii), there is some $N \geq l$ such that $M \vDash \mathrm{Pr}_N(\mathrm{Pr}_{l+1}(\mathrm{Pr}_l(\bot)) \to \mathrm{Pr}_s(\bot))$. W.l.o.g. pick this $N \geq n$. Since $N \geq n$, $M \vDash \mathrm{Pr}_N(\mathrm{Pr}_m(\mathrm{Pr}_l(\bot)))$, and therefore, $M \vDash \mathrm{Pr}_N(\mathrm{Pr}_s(\bot))$. Because $N > s$, we have $M \vDash \mathrm{Pr}_N(\mathrm{Pr}_{N-1}(\bot))$, which contradicts with (i), and the proof follows. □

1.9 A Remark on the Logic of Proofs

As we mentioned in the Introduction, and as far as we know, the only successful attempt to find a natural provability interpretation for **S4** and hence, a formalization of the BHK interpretation is done by Artemov [2] and is called the logic of proofs. In this section, we will look into this approach and investigate some of its advantages and disadvantages.

The main idea of the logic of proofs, **LP**, is using explicit proofs to avoid the non-standard proofs and hence to eliminate the incompleteness phenomenon. Let us give a more detailed account of this result. The language of **LP** is two sorted; one sort is for the explicit proofs and the other for the propositions. The first sort consists of proof terms constructed by the proof variables, proof constants and the proof connectives $+$, $\cdot$ and $!$, while the second sort contains terms constructed by the propositional variables, propositional connectives and the formulas of the form $t : A$ in which t is a proof term and A is a proposition. Let us explain the

intuitive meaning of these operations:

First of all we have to emphasize that in this interpretation, unlike the usual case in mathematics, proofs can be multi-conclusion. To find a natural candidate for these multi-conclusion proofs, it is enough to consider any usual proof as a proof for all intermediate statements it uses to prove the conclusion. For instance, the usual proof $A_1, A_2, \ldots, A_n$ of A_n will be interpreted as a proof for all A_i's.

1. The operation "!". If t is a proof for A, then $!t$ is a proof for the fact that "t is a proof for A". Therefore, the operator ! is the proof checker and could be interpreted as a self-awareness operator.
2. The operation "$\cdot$". If t is a proof for $A \to B$, and s is a proof for A, then $t \cdot s$ is a proof for B. Intuitively, $\cdot$ means the application of Modus Ponens on the proofs.
3. The operation "+". $t+s$ means the union of the proofs t and s. Recall that our proofs are multi-conclusion and $t+s$ can be served as a proof for all conclusions of t and s. Therefore if t is a proof for A and s is a proof for B, then $t+s$ is a proof for both A and B. To gain a better understanding, if we use the canonical way of changing usual proofs to multi-conclusion proofs, i.e. reading a usual proof as a proof for all intermediate statements in the proof, then $t+s$ just means putting t and s together. This is exactly what the symbol + suggests.
4. The predicate " : ". The intuitive meaning of $t : A$ is that t is a proof for A.

The formal system **LP** is a theory in this language to capture the intended meaning of the symbols defined above. The axioms are the following:

1. A finite complete set of axioms for the classical propositional logic for the language of **LP**,
2. $t : A \to A$,
3. $t : A \to B \to (s : A \to t \cdot s : B)$,
4. $t : A \to !t : t : A$,
5. $t : A \to s + t : A$,
6. $s : A \to s + t : A$.

The rules are the modus ponens and the neccesitation rule. The latter means that for any axiom A, we have $\vdash c_A : A$, where c_A is an appropriate constant exclusively used for A.

The natural interpretation for **LP** would be based on the usual proofs in Peano arithmetic. To formalize this idea, first of all we need a proof predicate: A proof predicate is a provably Δ_1 formula (in **PA**) $\mathrm{Prf}(x, y)$ with some natural basic properties (which we skip here. See [2]), and the following fundamental property:

$$\mathbf{PA} \vdash A \Leftrightarrow \exists x \mathrm{Prf}(x, \ulcorner A \urcorner).$$

We want to interpret the language of **LP** with this natural provability interpretation. Define an arithmetical substitution $*$ as the following: Firstly, it interprets $\cdot$, !, + and constants as the recursive functions on proofs in **PA** in the intended way. For instance, the function for $\cdot$ i.e., $\cdot^*$, will be the recursive function

which reads the codes of the proofs for A and $A \to B$ and replies the code of a proof for B. Why can we define such recursive functions? To show the fact that these functions exist, we need a proof; but here we just want to explain the main idea instead of a formal proof. For this reason, let us limit ourselves to the canonical proof predicate of **PA**. In this case, it can be easily shown that we can define these functions in a recursive way. For instance, if x and y are proofs for $A \to B$ and A respectively, for $\cdot^*(x, y)$ it is enough to put y after x and add the formula B at the end. This is obviously a proof for B and this process is clearly a recursive function. Moreover, note that for any c_A, c_A^* is one of the proofs for the axiom A^*. The existence of such a c_A^* also needs a proof, which we skip here. (See [2].)

Up to this point, we have interpreted all the proof connectives as recursive functions. Use these interpretations to interpret all proof terms t. Note that for interpreting proof variables we use arbitrary natural numbers as codes of proofs. Extend the interpretation $*$ to formulas. The idea is just interpreting all atoms as arithmetical sentences, reading $t : A$ as the proof predicate $\mathrm{Prf}(t^*, \ulcorner A^* \urcorner)$ and commute $*$ with all boolean connectives. For instance, the interpretation of $!x : p \to p$ would be $\mathrm{Prf}(!^*(n), \ulcorner \phi \urcorner) \to \phi$ where the interpretations of x and p are n and ϕ, respectively.

These arithmetical interpretations are the natural and concrete interpretations of the proofs, and in [2] Artemov proved that **LP** is sound and complete with respect to the class of these arithmetical interpretations.

Theorem 1.9.1. $\mathbf{LP} \vdash A$ *iff* A^* *is true for all arithmetical interpretations* $*$.

So far, we have found a natural proof interpretation for the system **LP**. Finding a natural interpretation for **S4** into **LP** would be the next step. Subsequently, we can use the composition of these interpretations to find a proof interpretation for **S4** and hence for **IPC**. We do not go into detail about the interpretation of the modal language into the system **LP**, but the basic idea is the following: Interpret any box as the existence of a proof; thus, any modal sentence will be equivalent to a first order formula in the language of **LP**. Therefore, we have quantifiers everywhere and specially in the scope of the predicate ":". We know that there is no way to exchange the quantifiers with the proof predicate (which is the reason why the incompleteness phenomenon and non-standard proofs appear), but since we require all the codes of the proofs to be standard numbers, we extract all the quantifiers and convert the translated formula into the prenex form. Use the Skolemization technique to witness the existential quantifiers by the universal ones. These witnesses are called realizations. (This is where we essentially need "+". It is important to note that by using Skolemization, we usually find a finite set of different witnesses and then we can roughly use + to merge these finite witnesses into one.) Note that this is not how Artemov argues in [2]; however, we explained the realizations in the way that we think is more accessible and to show why it is natural to have such a concept at the heart of the interpretation of the modal sentences. Let us illuminate the above interpretation by an example.

Example 1.9.2. *Consider the modal formula* $(\Box(p \to p) \wedge \neg\Box p) \to \Box\neg\Box p$. *First, we have to interpret all of the boxes as the existence of the proofs. Hence, we have*

$$(\exists w : (p \to p) \wedge \neg\exists x : p \to \exists y : (\neg\exists z : p).$$

Then, by extracting the quantifiers, we have

$$(\exists w : (p \to p) \wedge \forall x \neg x : p) \to \exists y \forall z\ y : \neg z : p,$$

which is equivalent to

$$\forall w \exists x \exists y \forall z((w : (p \to p) \wedge \neg x : p) \to y : \neg z : p)).$$

And finally by witnessing y *and* x *by some terms* $t(w, z)$ *and* $s(w, z)$, *we have*

$$(w : (p \to p) \wedge \neg s(w, z) : p) \to t(w, z) : \neg z : p.$$

This new formula is a realization for the modal formula $(\Box(p \to p) \wedge \neg\Box p) \to \Box\neg\Box p$. *Note that this realization is just one possible realization of the formula and if we change the witnessing terms* $t(w, z)$ *and* $s(w, z)$, *we can find different realizations for the same formula.*

After introducing the realizations, Artemov proved the following: (See [2].)

Theorem 1.9.3. $\mathbf{S4} \vdash A$ *iff there exists some realization* r *such that* $\mathbf{LP} \vdash A^r$.

In sum, we can say that Artemov used two ingredients to find a provability interpretation for **S4**. The first one is the interpretation of modal sentences via realizations into the system **LP**. (Here the main idea is the interpretation of the boxes as the existence of the standard proofs.) And the second ingredient is the interpretation of the system **LP** via natural arithmetical proof interpretations. Therefore, the main idea of what Artemov did, is to use the system **LP** as a bridge to interpret **S4** via arithmetical proof interpretations.

Let us explain the advantages of this approach. First of all, it uses the explicit proofs and by the method of using realizations, it makes sure that everything is a standard proof in this context. Therefore, this approach actually kills the effect of Gödel's incompleteness theorems and makes the proof interpretation more intuitive. Note that naturally, we do not count infinite non-standard proofs as proofs. Moreover, regardless of the relation between modal logics and explicit proofs, the system **LP** has its own applications. In fact, since it is a formal system for explicit proofs, it can be used as a theory to investigate the concept of proof and its natural calculus. Consequently, these formal systems are appropriate to investigate the formal verification in computer science or the behavior of justifications in formal epistemology.

However, this utopia of explicit proofs comes at a price. The price is a combination of two unintended properties: The first one is related to the fundamental change in the interpretation of the concept of provability and the second one is about the role of **LP** as an unbiased bridge. The problem is that the bridge is not neutral and somehow reflects its own behavior, which is not what we wanted.

Let us explain the first property by a simple example: Consider the modal sentence $\Box\neg\Box p$. The intended meaning of this sentence is the existence of a proof that shows p is not provable. In other words, it states that there *exists* a proof which shows that for *any* possible proof x for p, x is not a proof for p. Let us use the logic of proofs interpretation of the sentence. Since the occurrences of the inner and the outer box are negative and positive respectively, the meaning of the sentence is the existence of a term $t(x)$ such that $t(x) : \neg x : p$. Forgetting the condition that the term $t(x)$ should be a term in the language, it means that for all x, there exists a proof $y = t(x)$ which proves $\neg x : p$. In other words, it says that for *any* possible proof x for p, there *exists* a proof which shows that x is not a proof for p. It is easy to check that while the first interpretation is an $\exists\forall$ statement, the second one is a $\forall\exists$ statement, and it is obviously weaker than the first one. In fact, when we claim that we have a proof for unprovability of p, we mean a fixed uniform proof of the fact and we do not mean a machine (term) to transform a possible proof of x to a proof y that shows x is not a proof for p.
What we showed above is just the difference for one statement. Nevertheless, the argument actually works for different kinds of sentences. The reason is simple: Logic of proofs needs to kill the presence of non-standard numbers. For this matter, it pushes out all the quantifiers. (It also changes the order of quantifiers to find a functional interpretation of proofs.) Since quantifiers do not commute with proof predicates, the content of the sentence before pushing out the quantifiers is different from that of the transformed sentence. The first sentence is the intended interpretation of provability and the latter is what the logic of proofs interprets as the meaning of provability. While this new interpretation is interesting and useful, it is not the intended interpretation of informal provability and hence not *the* interpretation of **S4**.

In the following, we accept the functional interpretation of provability as what the logic of proofs proposed and we want to investigate the role of terms which we ignored in the previous argument. Let us explain the second property by a thought experiment: Think of the situation that you have another binary connective "?" in the language of **LP** with the following intuitive meaning: If s is a proof for $A \to A$ and t is not a proof for A, then $?(s,t)$ is a proof of the proposition that "t is not a proof for A". Add the axiom

$$(s : (A \to A) \wedge \neg t : A) \to ?(s,t) : \neg t : A$$

to the system **LP** and call it **LP**?. It is clear that the connective ? and the above sentence are the negative versions of the connective ! and its corresponding axiom, respectively. What is not clear is the use of the seemingly useless part $s : A \to A$. We can explain this issue as the following: Assume that we have a non-proof t for A and we want to construct a proof of the sentence $\neg t : A$. We call this proof r. The important fact is that the sole access to t is not enough to construct r because the code of A is also needed and this is actually where s plays its role: s is a proof for $A \to A$, hence we can use s to compute the code of A and now we have enough information to construct r.
Our method here seems ad-hoc and is certainly ugly, but remember that our goal is to perform an experiment about **LP** and fortunately this ad-hoc example is good enough to make our point. Now, let us be more formal about the natural

arithmetical interpretation of this connective and this new system. Since we used explicit standard proofs, we know that there exists a recursive function which reads t and the code of A and if t is not a proof for A, finds a proof of this fact. The reason is as follows: We know that $\mathrm{Prf}(x, y)$ is provably Δ_1, hence if $\neg\mathrm{Prf}(t, \ulcorner A \urcorner)$, we have

$$\mathbf{PA} \vdash \neg\mathrm{Prf}(t, \ulcorner A \urcorner).$$

Therefore, by the definition of a proof predicate we have

$$\exists r \mathrm{Prf}(r, \ulcorner \neg\mathrm{Prf}(t, \ulcorner A \urcorner) \urcorner).$$

Use unbounded search to find this r. Since it exists, our program halts and finds it. Now interpret $?(s, t)$ as the recursive function which reads s, finds the code of A and then by the above-mentioned method finds the intended proof r. Thus, based on this new natural arithmetical interpretation, we can interpret the new axiom $(s : (A \to A) \wedge \neg t : A) \to ?(s, t) : \neg t : A$. Hence, we have a natural arithmetical interpretation for the system **LP**?. On the other hand, one of the instances of the new axiom, i.e. $(w : (A \to A) \wedge \neg z : A) \to ?(w, z) : \neg z : A$, where z and w are proof variables, is the realization of the modal statement $\mathbf{5}' : (\Box(A \to A) \wedge \neg\Box A) \to \Box\neg\Box A$ in this new language. (Put $t(w, z) = ?(w, z)$ and $s(w, z) = z$ in Example 1.9.2.) The above discussion means that we can find a very natural provability interpretation of a variant of the axiom **5**. Recall that this axiom is not provable in **S4** and it seems contradictory with Artemov's completeness result. However, there is no contradiction. The reason is that "?" is not in the original language of **LP**, and hence you can not use it as a witness in the realization.

This observation shows that the arithmetical interpretation actually validates a variant of the axiom **5**, but the lack of the appropriate symbol in **LP** interferes with this fact. Therefore, the system **LP** does not reflect the whole power of explicit proofs; it just chooses the appropriate part to witness all the theorems of **S4** and nothing more than that. In other words, the formalization of the provability interpretation via the explicit proofs is very sensitive to the language we use. If we change the language, then with the same arithmetical interpretation, we will capture different modal logics. Therefore, we can conclude that the soundness-completeness result for **S4** with respect to this kind of arithmetical interpretations is a soundness-completeness result for the language we use and not the natural arithmetical interpretation we choose. Now, a natural question would be the following: If we eliminate this language barrier and make the relation between modal logics and arithmetical interpretations as "direct" as possible, then which modal logic corresponds to the whole power of the arithmetical interpretations of the proofs? By the direct connection, we roughly mean the following: For any modal sentence A, write it in the prenex form in a way that we defined before. Then, instead of witnessing the existential quantifiers by some terms in some language, witness them by some *natural recursive functions* on the proofs in Peano arithmetic. Define the logic E as the logic of all statements which are valid for this kind of arithmetical interpretations. Clearly, the question mentioned above is informal, but it is easy to verify that the answer is not **S4**. The reason is that we can find an appropriate way to interpret a variant of **5** as we have shown above. It is appropriate because there is no *a priori* reason to accept the recursive function ! and reject ?. The first one finds a proof for $\mathrm{Prf}(m, n)$ if $\mathrm{Prf}(m, n)$ is

true and the second function finds a proof for $\neg\mathrm{Prf}(m,n)$ if $\mathrm{Prf}(m,n)$ is false. Both of them are recursive and hence accessible for us as human beings. Note that Prf is a provably recursive predicate, and hence finding a proof for $\mathrm{Prf}(m,n)$ or a proof for its negation are similar computational tasks. (In the modal setting, the axioms **4** and **5** are intuitively different because we read $\Box A$ as $\exists x\mathrm{Prf}(x,A)$. This interpretation makes the sentence Σ_1 which is different from its negation.)

To sum up, the explicit proofs approach first kills all the quantifiers and puts some explicit witnesses for them. Therefore, it ignores the order of quantifiers and changes the canonical meaning of sentences and then as a consequence, it eliminates the computability based difference between provability and unprovability (Σ_1 vs Π_1) and maps both predicates to the boolean combinations of the explicit proof predicate Prf, which belongs to the class Δ_1. Consequently, the axioms **4** and **5** become similar and hence arithmetical interpretations can interpret a variant of **5** in a very natural way. Finally, to avoid this fact, the logic of proofs uses the language of **LP** to regain the difference between **4** and **5** by choosing what we need for **S4** and ignoring the other natural functions as exemplified by the function ?. This argument shows that the approach of explicit proofs does not distinguish **4** from **5** in a *natural* and *essential* way and hence, it can not be considered as a formalization of the provability interpretation of **S4**.

As the final part of this section, let us compare what we do in this chapter with the approach of the explicit proofs. First of all, we use the canonical meaning of provability instead of the logic of proofs' functional interpretation. Moreover, we do not use any language as a bridge. Therefore, our soundness-completeness results represent the provability behavior of our arithmetical interpretations in a direct way. Secondly, to capture different modal logics, we impose different natural conditions on our provability models, specifically on the hierarchy of theories. Therefore, we can claim that our approach can characterize different modal logics based on their different provability natures. Thirdly, our interpretation is based on the implicit proofs approach and hence it is a natural generalization of Solovay's work on **GL**. But since the Löb axiom is based on the incompleteness phenomenon, the explicit approach does not capture it and thus does not accept Solovay's provability interpretation as a special case. Hence, the explicit approach can not serve as the general framework for provability interpretations.

1.10 BHK Interpretations

Briefly, what we are going to do in this section, is to introduce a formalization of the BHK interpretation. Indeed, we will generalize this goal to make a framework to formalize different kinds of provability interpretations which includes the BHK interpretation as a special case. Note that the usual BHK interpretation is not the unique provability interpretation of the propositional language; in fact, there are many of them. Some of them can be characterized as variants of the original BHK interpretation, and some can't. The reason is that those provability interpretations do not satisfy the intended philosophical conditions which we want to have, but they are still provability interpretations and they need an exact formalization if we want to use them. Let us illuminate the idea by two examples.

The first one is a controversial variant of the BHK interpretation; it is obtained from the original BHK interpretation after relaxing the condition which says that there does not exist a proof for $\bot$. This interpretation informally corresponds to the minimal propositional logic, **MPC**. The second example of provability interpretation is also obtained from the original BHK interpretation, but now we read $\bot$ as the inconsistency, instead of the provability of the inconsistency. More precisely, and using the notation of Gödel's translation, we have $\bot^g = \bot$, where g stands for this new translation (which is different from what we used in the Introduction). This provability interpretation cannot be characterized as a variant of the BHK interpretation because of some philosophical reasons, which we do not get into here.

In this section, we try to justify the claim that our provability interpretation can offer an appropriate framework to formalize these different provability interpretations of the propositional logics. To implement this idea, we need two steps. First, we have to interpret all the connectives as the provability interpretation demands; this step is done by Gödel's translation. The second step is interpreting the provability predicates (i.e. boxes in the modal translation) as the classical provability of the classical theories. For that reason, we need a hierarchy of theories to formalize the hierarchy of the intuitive provabilities in the definition of the provability interpretation and also a model to evaluate the truth value of our statements. This second step is done by the provability models.

What we discussed above is the general framework. Let us come back to the specific case, which is the original BHK interpretation. Is there a *right* formalization of this interpretation? As we will show later, for different kinds of provability models, we have different BHK interpretations and these interpretations could show inherently different provability behaviors. Consequently, there are different formalizations for the BHK interpretation, instead of just a canonical one. The reason is that the BHK interpretation just interprets propositional connectives in a discourse of provability, but it does not say anything about the internal structure of the concept of provability. For instance, it does not say anything related to the power of the meta-theories compared to the lower theories. Since the BHK interpretation is the intended semantics for the intuitionistic logic, we have to accept that there could be different intuitionistic logics in terms of different interpretations of the power of our model and our theories. All of them are equally intuitionistic if we have just the BHK interpretation as the criterion.

The natural question is that what these intuitionistic logics are if we impose some natural conditions on the behavior of our model and our theories.
In the following, we will show that for some natural classes of provability models such as the class of all models or the class of all reflexive models, we can characterize some propositional logics such as **BPC** and **IPC**, respectively. For instance, in the case of reflexive models, the result shows that if we use the BHK interpretation with the philosophical commitment which states that all of the theories, meta-theories, meta-meta-theories and so on are sound and also, any meta-theory is powerful enough to prove the soundness of the lower theories, then the logic of the formulas which are valid under this kind of BHK interpreta-

tion, is the usual propositional intuitionistic logic. But, if we choose the minimal power, which does not assume any non-trivial condition on the hierarchy of the meta-theories, then the logic will change to **BPC**. However, what is important here is that all of these logics could be characterized as intuitionistic logics. This fact can explain the reason behind the disputes about finding the correct formalization of the intuitionistic logic. For instance, in [17], Ruitenburg argues that the *truly* intuitionistic logic is not **IPC** and he proposed **BPC** as the right one. Our approach here has a plural nature, and it tries to explain why with the same informal semantics (the BHK interpretation) there are different proposed logics.

Finally, a remark about classical logic. Since we have the axiom of the excluded middle in classical logic, we should have the following condition on provability models: *Either the "provability of p" is provable or it is provable that the provability of p implies the provability of $\bot$*. This means that the meta-theory should be powerful enough to prove the unprovability of almost all unprovable formulas. As we saw in the case of the logic **S5**, it contradicts with the natural condition that all the theories should be recursively enumerable. Therefore, intuitively speaking, we have to say that classical logic is beyond the scope of the BHK interpretation. In the following, we will prove this fact in a precise way.

Definition 1.10.1. *A provability interpretation for the propositional language is a translation from the propositional language to the language of modal logics.*

To illuminate the Definition 1.10.1, let us introduce three provability interpretations as examples.

Definition 1.10.2. *The BHK interpretation b is the following translation:*

(i) $p^b = \Box p$ *and* $\bot^b = \Box\bot$

(ii) $(A \wedge B)^b = A^b \wedge B^b$

(iii) $(A \vee B)^b = A^b \vee B^b$

(iv) $(A \to B)^b = \Box(A^b \to B^b)$

(v) $(\neg A)^b = \Box(A^b \to \Box\bot)$

Our translation is the same as the usual one, except for the case of $\bot$, which is translated to $\bot$ in the usual translation. (The negation of a formula A is considered as $A \to \bot$ and it inherits this change in the translation from $\bot$.) The reason for slightly changing the definition of the translation is because the usual translation can not capture the intended intuition of the BHK interpretation. Actually, the intended intuitionistic meaning of $\bot$, similar to the other atomic formulas, is its provability. Therefore, the natural interpretation of $\bot$ is $\Box\bot$. On the other hand, we know that the BHK interpretation claims that there is not any proof of $\bot$, which means $\neg\Box\bot$. Based on these two observations, we can justify the usual translation of $\bot$ as $\Box\bot \wedge \neg\Box\bot$, which is the same as $\bot$. Nevertheless, we have to emphasize that the condition of the unprovability of inconsistency is not related to the meaning of the connectives, and hence it should not interfere in the BHK interpretation; it is actually a commitment we impose on the discourse

of the provability. In our terms, the unprovability of the inconsistency asserts that the theories and meta-theories are consistent and it is obviously a property of the provability model and not a property of the connectives which we want to define. Hence, to formalize the original BHK interpretation, we need two ingredients; one is the b translation which is the formalization of the implicit BHK interpretation, and the second is the consistency condition on the provability models. The following definition formally states the second condition.

Definition 1.10.3. *A provability model $(M, \{T_n\}_{n=0}^{\infty})$ is called a BHK model if for any n, $M \vDash \neg \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$.*

Remark 1.10.4. *It seems that the natural consistency condition would be the consistency of all the theories. Yet, it is not enough. For instance, it is possible that all the theories in the hierarchy are consistent, but some meta-theory thinks that the lower theory is inconsistent, which contradicts with what an intuitionist assumes. For the intuitionist, the hierarchy of theories are just different layers of the story of the mind, and obviously these stories must be consistent in accordance with the BHK interpretation. However, this condition should be mentioned in the story itself. One way is assuming that any meta-theory actually proves the consistency of the lower theories. This is a natural condition, but it imposes a strong commitment on our theories. To keep the commitments as minimal as possible, we believe that the right condition to impose on the theories is the weaker condition which states that any meta-theory does not think that the lower theory is inconsistent. As we will see, this weaker condition widens the horizon of the BHK interpretation to capture the basic propositional logic on the one hand, and avoid artificial and degenerate models in which we could capture classical logic, on the other.*

Based on the aforementioned considerations, when we talk about the formalization of the BHK interpretation, we always refer to the BHK models. Let us formalize what we will call the weak BHK interpretation.

Definition 1.10.5. *Let q be a new atom which does not belong to the propositional language. The weak BHK interpretation, w, is the following translation:*

(i) $p^w = \Box p$ *and* $\bot^w = \Box q$

(ii) $(A \wedge B)^w = A^w \wedge B^w$

(iii) $(A \vee B)^w = A^w \vee B^w$

(iv) $(A \to B)^w = \Box(A^w \to B^w)$

(v) $(\neg A)^w = \Box(A^w \to \Box q)$

The translation is based on the idea that in this variant of the BHK interpretation, we eliminate the consistency condition from the discourse of provability. As a result, with this interpretation the intuitionist can not distinguish the inconsistency statement from any other statements. Therefore, in her viewpoint, $\bot$ is just a new atomic sentence which could be provable.

And finally, we will define Gödel's translation to show that there could be different provability models apart from the BHK interpretations.

Definition 1.10.6. *Gödel's provability interpretation, g, is the following translation:*

(i) $p^g = \Box p$ *and* $\bot^g = \bot$

(ii) $(A \wedge B)^g = A^g \wedge B^g$

(iii) $(A \vee B)^g = A^g \vee B^g$

(iv) $(A \to B)^g = \Box(A^g \to B^g)$

(v) $(\neg A)^g = \Box(\neg A^g)$

It is time to define the satisfaction of a propositional formula in a provability model with respect to some provability interpretation i.

Definition 1.10.7. *Let i be a provability interpretation. Then, by an expansion of a propositional formula A, and a witness for A under the interpretation i, we mean an expansion and a witness for A^i. And by $(M, \{T\}_{n=0}^{\infty}, i) \vDash \Gamma \Rightarrow A$ we mean $(M, \{T\}_{n=0}^{\infty}) \vDash \Gamma^i \Rightarrow A^i$. Moreover, if C is a class of provability models, by (C, i) we mean $\{(M, \{T\}_{n=0}^{\infty}, i) \mid (M, \{T\}_{n=0}^{\infty}) \in C\}$ and by $(C, i) \vDash \Gamma \Rightarrow A$ we mean $C \vDash \Gamma^i \Rightarrow A^i$.*

The next step is establishing the soundness-completeness theorem for the provability interpretations we defined. But first, we need a technical lemma.

Lemma 1.10.8. *If $\Gamma^b \vdash_{\mathbf{KD4}} A^b$, then* $\mathbf{EBPC} \vdash \Gamma \Rightarrow A$.

Proof. If $\Gamma^b \vdash_{\mathbf{KD4}} A^b$ then there is a cut-free proof for $\Gamma^b \Rightarrow A^b$ in $G(\mathbf{KD4})$. Call it π. It is clear that all formulas occurring in π are sub-formulas of A^b or sub-formulas of formulas in Γ^b. We know that all of these sub-formulas have the following forms: B^b; $B^b \to C^b$ and atoms p. ($\top$ and $\bot$ are considered atomic formulas in this proof.) Therefore, every sequent in π has the following form:

$$\Gamma^b, \{B_i^b \to C_i^b\}_{i \in I}, \{p_j\}_{j \in J} \Rightarrow \Delta^b, \{D_r^b \to E_r^b\}_{r \in R}, \{q_s\}_{s \in S}$$

Now we will prove the following claim:

Claim. If

$$G(\mathbf{KD4}) \vdash \ \Gamma^b, \{B_i^b \to C_i^b\}_{i \in I}, \{p_j\}_{j \in J} \Rightarrow \Delta^b, \{D_r^b \to E_r^b\}_{r \in R}, \{q_s\}_{s \in S}$$

where $\{p_j\}_{j \in J} \cap \{q_s\}_{s \in S} = \emptyset$ and $\bot \notin \{p_j\}_{j \in J}$ then for any $X \subseteq I$

$$\Gamma, \{D_r\}_{r \in R}, \{C_i\}_{i \in X} \vdash_{\mathbf{EBPC}} \bigvee\{\Delta, \{E_r\}_{r \in R}, \{B_i\}_{i \notin X}\}$$

The proof is by induction on the length of the cut-free proof in $G(\mathbf{KD4})$. To simplify the proof, we will call a sequent satisfying the conditions $\{p_j\}_{j \in J} \cap \{q_s\}_{s \in S} = \emptyset$ and $\bot \notin \{p_j\}_{j \in J}$, a good sequent.

The case for axioms and structural rules are easy to check. If the last rule is a conjunction or disjunction rule, then the main formula has the first form. Then since it is possible to simulate all conjunction and disjunction rules in **EBPC**,

the case of conjunction and disjunction rules are also easy to check. If the last rule is an implication rule, since we define our claim up to using implicational rules, there is nothing to prove in this case. Moreover, notice that if the conclusion sequent is good then the premises are so. Therefore, it is possible to use the induction hypothesis for them. Finally, if the last rule is a modal rule, then we have the following two cases:

1. If the last rule is a modal rule $\Box_4 R$, based on the form of formulas and the fact that in those three forms a boxed formula should be of the first kind, we have two cases. The first case is when the boxed formula in the right side has the form $\Box(D^b \to E^b)$. The second case is when the formula has the form $\Box p$. For the first case, the last rule has the following form:

$$\frac{\{p_j, \Box p_j\}_{j\in J}, \{B_i^b \to C_i^b, \Box(B_i^b \to C_i^b)\}_{i\in I} \Rightarrow D^b \to E^b}{\{\Box p_j\}_{j\in J}, \{\Box(B_i^b \to C_i^b)\}_{i\in I} \Rightarrow \Box(D^b \to E^b)}$$

and we want to prove

$$\{p_j\}_{j\in J}, \{B_i \to C_i\}_{i\in I} \vdash_{\mathbf{EBPC}} D \to E$$

Since every formula in the consequent sequent is boxed, it is a good sequent. Moreover, the only way for the premise sequent to not be good is that for some j, $p_j = \bot$. Therefore the claim is obvious from the $\bot$ rule in **EBPC**. Hence, we can also assume that the premise sequent is a good one. Then, by IH we know that for any $X \subseteq I$ we have

$$\{p_j\}_{j\in J}, \{B_i \to C_i\}_{i\in I}, \{C_i\}_{i\in X}, D \vdash_{\mathbf{EBPC}} \{B_i\}_{i\notin X}, E$$

By the rule $\to I$ the following is provable by $\Sigma = \{p_j\}_{j\in J} \cup \{B_i \to C_i\}_{i\in I}$

$$\bigwedge\{C_i\}_{i\in X} \wedge D \to \bigvee\{B_i\}_{i\notin X} \vee E$$

Fix $i \in I$ and also fix some $Z \subseteq I - \{i\}$. Both of the following statements are theorems of Σ:

$$\bigwedge\{C_i\}_{i\in Z} \wedge D \to \bigvee\{B_i\}_{i\notin Z} \vee \mathbf{B_i} \vee E$$

and

$$\bigwedge\{C_i\}_{i\in Z} \wedge \mathbf{C_i} \wedge D \to \bigvee\{B_i\}_{i\notin Z} \vee E$$

Since $\Sigma \vdash \mathbf{B_i} \to \mathbf{C_i}$. Then by using appropriate formalized rules we will have

$$\bigwedge\{C_i\}_{i\in Z} \wedge D \to \bigvee\{B_i\}_{i\notin Z} \vee E$$

provable by Σ in **EBPC**. By iterating this method we can eliminate all elements in I. Therefore we will have

$$\Sigma \vdash_{\mathbf{EBPC}} D \to E$$

which is what we wanted to prove.

If the boxed formula in the right side of the rule is $\Box p$, then the last rule has the form

$$\frac{\{p_j, \Box p_j\}_{j\in J}, \{B_i^b \to C_i^b, \Box(B_i^b \to C_i^b)\}_{i\in I} \Rightarrow p}{\{\Box p_j\}_{j\in J}, \{\Box(B_i^b \to C_i^b)\}_{i\in I} \Rightarrow \Box p}$$

and we want to prove

$$\{p_j\}_{j\in J}, \{B_i \to C_i\}_{i\in I} \vdash_{\mathbf{EBPC}} p$$

There are two different cases. The first case is when $p \in \{p_j\}_{j\in J}$ or $\bot \in \{p_j\}_{j\in J}$. In this case the claim is an obvious consequence of an axiom in **EBPC**. The second case is when $p \notin \{p_j\}_{j\in J}$ and $\bot \notin \{p_j\}_{j\in J}$. Therefore, the premise sequent is a good one. Hence by IH and for any $X \subseteq I$ we have

$$\{p_j\}_{j\in J}, \{B_i \to C_i\}_{i\in I}, \{C_i\}_{i\in X} \vdash_{\mathbf{EBPC}} \{B_i\}_{i\notin X}$$

with the same method as above we can deduce

$$\{p_j\}_{j\in J}, \{B_i \to C_i\}_{i\in I} \vdash_{\mathbf{EBPC}} \top \to \bot$$

Then by the rule C, we will have

$$\{p_j\}_{j\in J}, \{B_i \to C_i\}_{i\in I} \vdash_{\mathbf{EBPC}} \bot$$

which is what we wanted.

2. If the last rule is $\Box_D R$, then everything in the proof is the same as the proof for the case 1 when we put $D = \top$ and $E = \bot$. Therefore, we will have

$$\{p_j\}_{j\in J}, \{B_i \to C_i\}_{i\in I} \vdash_{\mathbf{EBPC}} \top \to \bot$$

Then by the rule C, we will have

$$\{p_j\}_{j\in J}, \{B_i \to C_i\}_{i\in I} \vdash_{\mathbf{EBPC}} \bot$$

which is what we wanted.
After proving the claim, the theorem is an easy consequences of the claim. Since there is a proof of $\Gamma^b \Rightarrow A^b$ in $G(\mathbf{KD4})$ then the sequent is obviously a good one and hence by the claim we will have $\Gamma \vdash_{\mathbf{EBPC}} A$.

□

Theorem 1.10.9. (*i*) $\Gamma \vdash_{\mathbf{BPC}} A$ *iff* $\Gamma^b \vdash_{\mathbf{K4}} A^b$

(*ii*) $\Gamma \vdash_{\mathbf{EBPC}} A$ *iff* $\Gamma^b \vdash_{\mathbf{KD4}} A^b$

(*iii*) $\Gamma \vdash_{\mathbf{IPC}} A$ *iff* $\Gamma^b \vdash_{\mathbf{S4}} A^b$

(*iv*) $\Gamma \vdash_{\mathbf{FPL}} A$ *iff* $\Gamma^b \vdash_{\mathbf{GL}} A^b$

(*v*) $\Gamma \vdash_{\mathbf{MPC}} A$ *iff* $\Gamma^w \vdash_{\mathbf{S4}} A^w$

Proof. The proof of the soundness part is easy and routine. For the completeness part, the case (*iv*) is proved by Visser in [22]. The same proof also works for (*i*). (*iii*) is a well-known result. (See [18] for instance.) (*ii*) is proved by Lemma 1.10.8. For the case (*v*), we know that **MPC** and **S4** are sound and strongly complete with respect to the class of reflexive transitive Kripke models. (For

MPC the model should also be persistent.) However, in the case of **MPC**, the nodes can also satisfy $\bot$. Soundness is again easy. For the completeness part, if we have a counter **MPC**-Kripke model for $\Gamma \Rightarrow A$, we can construct a counter **S4**-model for $\Gamma^w \Rightarrow A^w$ in the following way: Use the same Kripke model, with the same values, but assume that q is true in a node, if $\bot$ is true in that node. Then, it is easy to show that for any propositional formula B, B is true in the node l iff B^w is so. Therefore, if the first model is a counter example for $\Gamma \Rightarrow A$, then the new one is a counter example for $\Gamma^w \Rightarrow A^w$. This construction proves the completeness part. □

We can use the soundness and completeness of these translations to transfer our results from the modal setting to the propositional one.

Definition 1.10.10. *The class* **BHK** *is the class of all BHK models and the class* **cBHK** *is the class of all BHK models which are constant.*

Theorem 1.10.11. (*i*) $\Gamma \vdash_{\mathbf{BPC}} A$ *iff* $(\mathbf{PrM}, b) \vDash \Gamma \Rightarrow A$. *And* $\mathbf{BPC} \vdash A$ *iff* $(\mathbf{BHK}, b) \vDash A$.

(*ii*) $\Gamma \vdash_{\mathbf{EBPC}} A$ *iff* $(\mathbf{Cons}, b) \vDash \Gamma \Rightarrow A$.

(*iii*) $\Gamma \vdash_{\mathbf{IPC}} A$ *iff* $(\mathbf{Ref}, b) \vDash \Gamma \Rightarrow A$.

(*iv*) $\Gamma \vdash_{\mathbf{FPL}} A$ *iff* $(\mathbf{Cst}, b) \vDash \Gamma \Rightarrow A$. *And* $\mathbf{FPL} \vdash A$ *iff* $(\mathbf{cBHK}, b) \vDash A$.

(*v*) *Let* $(M, \{T_n\}_{n=0}^{\infty})$ *be a provability model. Then* $(M, \{T_n\}_{n=0}^{\infty}, b) \vDash \mathbf{CPC}$ *iff there exists* n *such that* $M \vDash \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$. *Therefore, no BHK interpretation for classical logic exists.*

Proof. Based on Theorem 1.10.9, the strong soundness-completeness parts are just easy consequences of the soundness-completeness results for the corresponding modal logics. For the BHK completeness part for (*i*), if $(\mathbf{BHK}, b) \vDash A$, then there are expansions B_i's for A^w and a witness for $\bigvee B_i$, such that for all arithmetical substitutions σ, and all BHK models $(M, \{T_n\}_{n=0}^{\infty})$, we have $M \vDash (\bigvee_{i=0}^{r} B_i)^{\sigma}(w)$. Let Γ be a sequence of infinitely many copies of $\neg\Box\Box\bot$ and u a witness, which witnesses each of these formulas by $(n+1, n)$. We claim that for any provability model $(M, \{T_n\}_{n=0}^{\infty})$ and any arithmetical substitution σ, we have $M \vDash \Gamma^{\sigma}(u) \Rightarrow (\bigvee_{i=0}^{r} B_i)^{\sigma}(w)$. If $M \vDash \Gamma^{\sigma}(u)$, then for any n, we have $M \vDash \neg\mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$. Hence, $(M, \{T_n\}_{n=0}^{\infty})$ is a BHK model and therefore, $M \vDash (\bigvee_{i=0}^{r} B_i)^{\sigma}(w)$. We know $\mathbf{PrM} \vDash \Gamma \Rightarrow A^b$; therefore, by strong completeness for **K4**, we have $\Gamma \vdash_{\mathbf{K4}} A^b$. Thus, $\mathbf{K4} \vdash \neg\Box\Box\bot \to A^b$ and then, $\mathbf{K4} \vdash ((\top \to \bot) \vee A)^b$. By Theorem 1.10.9, $\mathbf{BPC} \vdash (\top \to \bot) \vee A$, and therefore by the disjunction property of **BPC**, we know that $\mathbf{BPC} \vdash A$ or $\mathbf{BPC} \vdash \top \to \bot$. The latter is impossible by simple facts about **BPC**, therefore $\mathbf{BPC} \vdash A$.
The case (*iv*) also needs an argument exactly similar to the case (*i*). Moreover, since the consistent and reflexive models satisfy the consistency condition of the BHK interpretation, the cases (*ii*) and (*iii*) are just a combination of Theorem 1.10.9 and the completeness results for the corresponding theories.

For (*v*) we need some justification. First of all we want to show that if for any n, $M \vDash \neg\mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$, then $(M, \{T_n\}_{n=0}^{\infty})$ is not a model for **CPC**. We prove

this claim by contradiction. Assume that for any n, $M \vDash \neg \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$ and $(M, \{T_n\}_{n=0}^{\infty}, b) \vDash \mathbf{CPC}$. We want to show that all three statements of the proof of Theorem 1.8.1 are also true in our case. Firstly, (i) is true by assumption. Secondly, consider the formula $\Box^n \top$ which is a translation of the propositional classical theorem $\top^n$ with the definition $\top^0 = \top$ and $\top^{n+1} = \top \to \top^n$. Therefore, the formula $\Box(\Box\Box(\Box\bot \wedge \Box^n\top) \to \Box\bot)$ is the translation of the tautology $((\top \to (\top \to (\bot \wedge \top^n))) \to \bot)$. Thus,

$$(M, \{T_n\}_{n=0}^{\infty}) \vDash \Box(\Box\Box(\Box\bot \wedge \Box^n\top) \to \Box\bot).$$

Since we used this formula to show (ii), we can claim that we also have (ii) here. Thirdly, we know that $p \vee \neg p$ is a theorem of **CPC**. Hence, $(M, \{T_n\}_{n=0}^{\infty}) \vDash (p \vee \neg p)^b$, which means $(M, \{T_n\}_{n=0}^{\infty}) \vDash (\Box p \vee \Box(\Box p \to \Box\bot)$. Therefore, (iii) is also true in M. Thus, we have a contradiction and it proves the claim.

For the converse, assume that there is some n such that $M \vDash \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$; we will show that $(M, \{T_n\}_{n=0}^{\infty}, w) \vDash \mathbf{CPC}$. First of all, to simplify the proof, define the complexity of any box as the maximum depth of the nested boxes in front of that box. For instance, the complexity of the inner box in $\Box(\Box p \wedge q)$ is zero, and the complexity of the outer box is one. Define the canonical witness starting from n, as follows: Witness any box by its complexity plus n. It is easy to show that this witness is an ordered one, because the witness for any outer box is bigger than the witness for the inner boxes. Define A^σ as the formula resulted by substituting all the atoms by σ and witnessing all the boxes by the canonical witness starting from n. It is easy to verify that for any propositional formula $A \to B$, $M \vDash ((A \to B)^w)^\sigma$. To show this, firstly, note that the following claim holds: For any propositional formula B,

$$I\Sigma_1 \vdash \mathrm{Pr}_n(\bot) \to (B^w)^\sigma.$$

The proof of the claim is based on induction on B and easily follows. Assume that the complexity of the outmost box in $\Box(A^w \to B^w)$ is $k \geq n+1$. (Since witnesses begin with n and there is at least one box in A^w, k is at least $n+1$.) By Σ_1-completeness we have

$$I\Sigma_1 \vdash \mathrm{Pr}_k(\mathrm{Pr}_n(\bot) \to (B^w)^\sigma),$$

and hence,

$$I\Sigma_1 \vdash \mathrm{Pr}_k(\mathrm{Pr}_n(\bot)) \to \mathrm{Pr}_k((B^w)^\sigma).$$

Then since $M \vDash I\Sigma_1$, then

$$M \vDash \mathrm{Pr}_k(\mathrm{Pr}_n(\bot)) \to \mathrm{Pr}_k((B^w)^\sigma).$$

We know that $M \vDash \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$ and $k \geq n+1$; hence $M \vDash \mathrm{Pr}_k(\mathrm{Pr}_n(\bot))$. Therefore,

$$M \vDash \mathrm{Pr}_k((B^w)^\sigma),$$

and thus,

$$M \vDash \mathrm{Pr}_k((A^w)^\sigma \to (B^w)^\sigma),$$

and the proof follows.

It is easy to check that for any formula B, there exists another formula C such that C is in the CNF form, in which all the literals are implicational formulas, positive atoms and $\bot$ and classically equivalent to B. Note that the process of constructing this C just uses the classical rules for conjunction and disjunction. Since w and the canonical witness respect the conjunction and disjunction and their basic rules, $(B^w)^\sigma$ and $(C^w)^\sigma$ are equivalent in M. Suppose that $\mathbf{CPC} \vdash B$; we want to show that $M \vDash (B^w)^\sigma$. It is enough to show that $M \vDash (C^w)^\sigma$. Considering that all the literals in C are implicational formulas, positive atoms and $\bot$, the literals of C^b are translations of implications, boxed atoms or $\Box\bot$. If $M \nvDash (C^w)^\sigma$, there must be some clause in which all the literals are false. Since the translations of the implications are true in M, there has to be a clause in C consisting of atoms and $\bot$. Therefore, C can not be a classical tautology and hence B will not be, as well. But $\mathbf{CPC} \vdash B$; a contradiction. Thus, $M \vDash (B^w)^\sigma$. So far, we have shown that if $\mathbf{CPC} \vdash B$, then $M \vDash (B^w)^\sigma$. If we send q in the definition of $\bot^w = \Box q$, to $\bot$, then we have $M \vDash (B^b)^\sigma$, which proves the theorem. $\Box$

There is another type of the BHK interpretation in which there is not any kind of assumption on the non-existence of a proof of the contradiction.

Theorem 1.10.12. *(i)* $\Gamma \vdash_{\mathbf{MPC}} A$ *iff* $(\mathbf{Ref}, w) \vDash \Gamma \Rightarrow A$.

(ii) Let $(M, \{T_n\}_{n=0}^\infty)$ *be a provability model. Then* $(M, \{T_n\}_{n=0}^\infty, w) \vDash \mathbf{IPC}$ *iff* $(M, \{T_n\}_{n=0}^\infty, w) \vDash \mathbf{CPC}$ *iff there exists* n *such that* $M \vDash \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$.

Proof. For (i), use Theorem 1.10.9 and the soundness-completeness results for **S4**. For (ii), if there exists n such that $M \vDash \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$, then by the proof of Theorem 1.10.11 part (v), we know that $(M, \{T_n\}_{n=0}^\infty, w) \vDash \mathbf{CPC}$. Moreover, if $(M, \{T_n\}_{n=0}^\infty, w) \vDash \mathbf{CPC}$, then we can easily verify that we have $(M, \{T_n\}_{n=0}^\infty, w) \vDash \mathbf{IPC}$. It remains to show that if $(M, \{T_n\}_{n=0}^\infty, w) \vDash \mathbf{IPC}$, then there exists n such that $M \vDash \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$.
Assume that $(M, \{T_n\}_{n=0}^\infty, w) \vDash \mathbf{IPC}$ and for any n, $M \vDash \neg\mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$. We want to reach a contradiction. We know that $\mathbf{IPC} \vdash \bot \to p$. Hence, $(M, \{T_n\}_{n=0}^\infty) \vDash (\bot \to p)^w$. Thus, $(M, \{T_n\}_{n=0}^\infty) \vDash \Box(\Box q \to \Box p)$. Consequently, there are expansions of the form, $\Box(\bigvee_{j=0}^{s_i}(\Box q \to \Box p))$ for $0 \leq i \leq r$ and witnesses $w_i = (n_i, (m_{ij}, k_{ij})_{j=0}^{s_i})$ such that for any arithmetical substitution σ,

$$M \vDash \bigvee_{i=0}^{r} \Box(\bigvee_{j=0}^{s_i}(\Box q \to \Box p))^\sigma(w_i).$$

Define $k = max_{ij}(k_{ij})$, $m = min_{ij}(m_{ij})$ and $n = max_i(n_i)$. It is easy to see that

$$M \vDash \mathrm{Pr}_n((\mathrm{Pr}_m(q^\sigma) \to \mathrm{Pr}_k(p^\sigma))).$$

And if we choose a substitution σ such that $q^\sigma = (0 = 0)$ and $p^\sigma = (0 = 1)$, then we have

$$M \vDash \mathrm{Pr}_n((\mathrm{Pr}_m(0 = 0) \to \mathrm{Pr}_k(0 = 1)),$$

and hence $M \vDash \mathrm{Pr}_n(\mathrm{Pr}_k(\bot))$. Thus, for some number $N > n, k$, we have $M \vDash \mathrm{Pr}_{N+1}(\mathrm{Pr}_N(\bot))$ which is a contradiction. $\Box$

2. Computational Flows in Arithmetic

2.1 Introduction

Intuitively speaking, proofs are the information carriers that transfer the informational content of the assumptions to the informational content of the conclusion. This open notion of content admits many different interpretations in many different disciplines. The most trivial one is the truth value which is preserved along any sound proof and consequently is the least informative one. But there are more useful examples. *The computational content* is one of them and it is no exaggeration to state that this type of content is one of the main players in proof theory and theoretical computer science. The reason is its widespread incarnations, from witnesses of existential quantifiers a la Herbrand to Gödel's Dialectica interpretation of higher order arithmetical statements. In this chapter, we will follow this line to introduce another computational interpretation which can be seen as a classical and a more direct reading of the Dialectica interpretation and in the rest of this introduction we will try to explain its basic ideas.

Let us begin with the general idea of how any computational interpretation works by summarizing the process behind it: First, any interpretation needs to interpret a sentence as a computational problem for which the computational content roughly means *any way* that can solve the problem computationally. Then it should define a computational flow as a sequence of certain types of *simple methods* to transfer the previously defined content from one point to another. And finally, it should find a way to translate any formal proof of a given system to such a computational flow.

To implement these three stages, we first need to define the game theoretic interpretation of formulas in the prenex form. The basic idea is the following: First interpret any quantifier-free formula $A(x_1, y_1, x_2, \ldots)$ as a game between two players in which the first player, plays x_1 and then the second player, plays y_1 and they continue this process alternately. At the end of the game, if $A(x_1, y_1, x_2, \ldots)$ holds, the second player wins and otherwise the first one is the winner. Now, the sentence $\forall x_1 \exists y_1 \forall x_2 \ldots A(x_1, y_1, x_2, \ldots)$ simply means that the second player has a winning strategy and this strategy is exactly the *computational content* of the sentence $\forall x_1 \exists y_1 \forall x_2 \ldots A(x_1, y_1, x_2, \ldots)$. Now let us define the *simple methods* or the *ways that the information flows*. In this stage, we have essentially two different choices to make. Let us first begin with the simpler one, i.e., the deterministic reductions. To explain these reductions, assume that we have a second player's strategy to win the game $A(x_1, y_1, x_2, \ldots)$ and we want to provide a second player's strategy to win the game $B(u_1, v_1, u_2, \ldots)$. For this purpose, we define the deterministic reduction from $\forall u_1 \exists v_1 \forall u_2 \ldots B(u_1, v_1, u_2, \ldots)$ to $\forall x_1 \exists y_1 \forall x_2 \ldots A(x_1, y_1, x_2, \ldots)$ as a tuple of functions (f_i, g_i) with the lowest possible complexity such that f_i reads all u_j's for $j \leq i$ and y_k for $k < i$ and finds

x_i, and g_i reads the same data plus y_i and computes v_i such that

$$A(f_1(u_1), y_1, f_2(u_1, u_2, y_1), \ldots) \to B(u_1, g_1(u_1, y_1), u_2, \ldots).$$

It is clear that these functions find a way to relate the moves of the games A and B to transfer the winning strategy of the second player for the game A to his winning strategy for the game B.

Now let us review the second type of reductions, i.e., the non-deterministic reductions. The idea behind the non-deterministic reductions is simply the power to first compute a move in a given stage and then after a while coming back to that move again to compute another value for the same move. Let us explain the idea by some examples. Assume that in the first step of the reduction, when we read u_1 to compute the move x_1, instead of using just one function f_1, we use two functions f_1 and f_1' to compute two possible choices for x_1's. Then these two choices lead to two possibly different moves y_1 and y_1' which we can use both to find our needed move v_1. If we continue with the usual deterministic reductions, this leads to the condition that states the formula

$$A(f_1(u_1), y_1, f_2(u_1, u_2, y_1, y_1'), \ldots) \wedge A(f_1'(u_1), y_1', f_2'(u_1, u_2, y_1, y_1'), \ldots)$$

should imply the formula $B(u_1, g_1(u_1, y_1, y_1'), u_2, \ldots)$. This is the simplest way of using non-determinism, but there are other more complex ways to consider. For instance, we can first use $f_1(u_1)$ to compute x_1 and then after reading y_1, we can again try to compute x_1, this time using the function $f_1''(u_1, y_1)$ which also has access to y_1. As we can observe, non-determinism can easily lead to some sort of hugely complex interaction between the moves which makes the transferring process for the winning strategies extremely complicated. We will investigate these reductions in the following sections.

So far, we have completed the definition of our reasonable methods. But what about the simplicity of these methods? At first glance, it seems that the low complexity of the reductions ensures the expected simplicity that we promised but unfortunately the reality is far from that. In fact, in some cases, while the complexity of the functions can be extremely low, verifying the truth of the formula

$$A(f_1(u_1), y_1, f_2(u_1, u_2, y_1), \ldots) \to B(u_1, g_1(u_1, y_1), u_2, \ldots)$$

in the deterministic case or the corresponding formula in the non-deterministic case, can be extremely high, non-trivial and non-syntactical. This is clearly not what we expect from a simple reduction. Hence, we also add a base weak theory $\mathcal{B}$ to the definition to force the above-mentioned implications to be provable in $\mathcal{B}$. This condition makes the reductions simple and syntactical as we expect them to be.

Based on these reductions, it is now natural to define a computational flow as a uniform sequence of reductions and try to transform any proof in any appropriate theory to a computational flow. This completes all the needed ingredients for our new proof mining technique to characterize the low complexity consequences of both weak and strong theories of arithmetic. The rest of this chapter is devoted to develop the details of this new proof mining method.

2.2 Preliminaries

In this section we will review some preliminaries. First, let us fix a language which can be any arbitrary extension of a ring-type language for numbers:

Definition 2.2.1. *Let $\mathcal{L}$ be a first order language of arithmetic extending $\mathcal{L}_{\mathcal{R}} = \{0, 1, +, \dot{-}, \cdot, d(-,-), \leq\}$ where $x \dot{-} y$ and $d(x, y)$ mean $max\{0, x - y\}$ and $\lfloor \frac{x}{y+1} \rfloor$ in the standard model, respectively. By $\mathcal{R}$ we mean the first order theory in the language $\mathcal{L}_{\mathcal{R}}$ consisting of the axioms of non-trivial commutative discrete ordered semirings (the usual axioms of non-trivial commutative rings minus the existence of additive inverse, plus the axioms to state that $\leq$ is a total discrete order such that $<$ is compatible with addition and multiplication with non-zero elements), plus the following defining axioms for $\dot{-}$ and d:*

$$(x \geq y \to (x \dot{-} y) + y = x) \wedge (x < y \to x \dot{-} y = 0),$$

and

$$((y + 1) \cdot d(x, y) \leq x) \wedge (x \dot{-} (y + 1) \cdot d(x, y) < y + 1).$$

Note that to avoid division by zero and to have a total function symbol in the language we defined division as $\lfloor \frac{x}{y+1} \rfloor$ and not $\lfloor \frac{x}{y} \rfloor$.

Remark 2.2.2. *First note that $\mathcal{R}$ can prove that all elements are non-negative simply because multiplying them preserves the order. Secondly note that the language $\mathcal{L}$ is powerful enough to represent the conditional function*

$$C(x, y, z) = \begin{cases} y & x = 0 \\ z & x > 0 \end{cases}$$

as a term and $\mathcal{R}$ is powerful enough to prove that the term works. The crucial point is that the term $\chi_{=0}(x) = d(x + 2, x) \dot{-} 1 = \lfloor \frac{x+2}{x+1} \rfloor \dot{-} 1$ has the following property provably in $\mathcal{R}$:

$$\chi_{=0}(x) = \begin{cases} 1 & x = 0 \\ 0 & x > 0 \end{cases}$$

Hence it is enough to represent C by $\chi(x)y + (1 \dot{-} \chi(x))z$. Moreover, using $\chi_{\leq}(x, y) = \chi_{=0}(x \dot{-} y)$, we can represent the characteristic function for $\leq$ and since we have the power to simulate all boolean operators and $x = y$ is equivalent to $x \leq y \wedge y \leq x$, we have the characteristic functions of all quantifier-free formulas of the language $\mathcal{L}_{\mathcal{R}} = \{0, 1, +, \dot{-}, \cdot, d(-,-), \leq\}$.

To define different bounded systems of arithmetic, we have to set two main ingredients of the induction axiom, i.e., the complexity of the induction formula and the length of the induction. For the first one we have:

Definition 2.2.3. *The hierarchy $\{\Sigma_k, \Pi_k\}_{k=0}^{\infty}$ is defined recursively in the following way:*

(i) $\Pi_0 = \Sigma_0$ *is the class of all quantifier-free formulas,*

(ii) $\Sigma_k \subseteq \Sigma_{k+1}$ *and* $\Pi_k \subseteq \Pi_{k+1}$,

(*iii*) Π_k *and* Σ_k *are closed under conjunction and disjunction,*

(*iv*) *If* $B(x) \in \Sigma_k$ *then* $\exists x \leq t\ B(x) \in \Sigma_k$ *and* $\forall x \leq t\ B(x) \in \Pi_{k+1}$ *and*

(*v*) *If* $B(x) \in \Pi_k$ *then* $\forall x \leq t\ B(x) \in \Pi_k$ *and* $\exists x \leq t\ B(x) \in \Sigma_{k+1}$.

Remark 2.2.4. *In this chapter,* Π_k *and* Σ_k *always mean the previously defined classes of bounded formulas. To denote the usual arithmetical hierarchy based on unbounded quantifiers we will use* Π^0_k *and* Σ^0_k. *Moreover, whenever* A *is quantifier-free and* $B \in \bigcup_{k\geq 0} \Sigma_k$, *we use the formula* $A \to B$ *as an abbreviation for* $\neg A \vee B$. *Unlike the former, the latter is a formula in* $\bigcup_{k\geq 0} \Sigma_k$.

Definition 2.2.5. (*i*) *A class of bounded formulas* Π *is called a* π*-class of the language* $\mathcal{L}$ *if it includes all quantifier-free formulas of* $\mathcal{L}$, *is closed under substitutions, subformulas, conjunction, disjunction and bounded universal quantifiers and if* $\exists y \leq t\ B(y) \in \Pi$ *then there exists* $C(y)$ *such that* $\vdash C(y) \leftrightarrow \neg B(y)$ *and* $\forall y \leq t\ C(y) \in \Pi$.

(*ii*) *A class of bounded formulas* Σ *is called a* σ*-class of the language* $\mathcal{L}$ *if it includes all quantifier-free formulas of* $\mathcal{L}$, *is closed under substitutions, subformulas, conjunction, disjunction and bounded existential quantifiers and if* $\forall y \leq t\ B(y) \in \Sigma$ *then there exists* $C(y)$ *such that* $\vdash C(y) \leftrightarrow \neg B(y)$ *and* $\exists y \leq t\ C(y) \in \Sigma$.

Example 2.2.6. *The class of all bounded formulas is a trivial example of both* π *and* σ *classes. The more interesting examples though include the classes* $\hat{\Pi}^b_k(\#_m)$ *and* $\hat{\Sigma}^b_k(\#_m)$ *(drop* $\#_m$ *from the notation when* $m = 2$*), in the language of bounded arithmetic augmented with subtraction, division and* $\#_i$ *for* $2 \leq i \leq m$. *These classes are defined in the following way:*

(*i*) $\hat{\Pi}^b_0(\#_m) = \hat{\Sigma}^b_0(\#_m)$ *is the class of all sharply bounded formulas, i.e., the formulas whose quantifiers are bounded by a term of the form* $|t|$, *for some term* t,

(*ii*) $\hat{\Sigma}^b_k(\#_m) \subseteq \hat{\Sigma}^b_{k+1}(\#_m)$ *and* $\hat{\Pi}^b_k(\#_m) \subseteq \hat{\Pi}^b_{k+1}(\#_m)$,

(*iii*) $\hat{\Pi}^b_k(\#_m)$ *and* $\hat{\Sigma}^b_k(\#_m)$ *are closed under conjunction and disjunction,*

(*iv*) *If* $B(x) \in \hat{\Sigma}^b_k(\#_m)$ *then* $\exists x \leq t\ B(x) \in \hat{\Sigma}^b_k(\#_m)$ *and* $\forall x \leq t\ B(x) \in \hat{\Pi}^b_{k+1}(\#_m)$ *and*

(*v*) *If* $B(x) \in \hat{\Pi}^b_k(\#_m)$ *then* $\forall x \leq t\ B(x) \in \hat{\Pi}^b_k(\#_m)$ *and* $\exists x \leq t\ B(x) \in \hat{\Sigma}^b_{k+1}(\#_m)$.

We can also consider a more relaxed version of these classes, i.e., $\Sigma^b_k(\#_m)$ *and* $\Pi^b_k(\#_m)$, *(again dropping* $\#_m$ *when* $m = 2$*), which are defined with the same definition as above, adding the condition that:*

"$\Pi^b_k(\#_m)$ *and* $\Sigma^b_k(\#_m)$ *are closed under sharply bounded quantification, i.e., a quantification bounded by* $|t|$ *for some term* t*."*

Note that, assuming that the polynomial hierarchy does not collapse, these more relaxed versions of the classes (for $m = 2$) are not π- and σ-classes, respectively. The reason is the existence of a Π_k^b formula (a Σ_k^b formula), ending with an existential (a universal) sharply bounded quantifier, which is also bounded, without a Π_k^b negation (a Σ_k^b negation).

Now let us define a robust form for the classes of terms that can play the role of induction-length.

Definition 2.2.7. *Let $\mathcal{A} \supseteq \mathcal{R}$ be a theory. A class of terms, $\mathbb{T}$, is called an $\mathcal{A}$-term ideal if:*

(i) It is closed under all function symbols of the language $\mathcal{L}_\mathcal{R}$, provably in $\mathcal{A}$, i.e. for any function symbol $f \in \mathcal{L}_\mathcal{R}$ and any $t(\vec{x}) \in \mathbb{T}$, there exist $r(\vec{x}) \in \mathbb{T}$ such that $\mathcal{A} \vdash r(\vec{x}) = f(t(\vec{x}))$.

(ii) It is closed under substitution, i.e. if $t(\vec{x}, y) \in \mathbb{T}$ and s is an arbitrary term (not necessarily in $\mathbb{T}$) then $t(\vec{x}, s) \in \mathbb{T}$ provably in $\mathcal{A}$, i.e. there exists $r(\vec{x}) \in \mathbb{T}$ such that $\mathcal{A} \vdash r(\vec{x}) = t(\vec{x}, s)$.

(iii) It has a subset of monotone majorizing terms provably in $\mathcal{A}$, i.e. there exists a set of terms $M \subseteq \mathbb{T}$ such that for any $t(\vec{x}) \in \mathbb{T}$ there exists $s(\vec{x}) \in M$ such that $\mathcal{A} \vdash t(\vec{x}) \leq s(\vec{x})$ and for any $r(\vec{x}) \in M$, $\mathcal{A} \vdash \vec{x} \leq \vec{y} \rightarrow r(\vec{x}) \leq r(\vec{y})$.

Example 2.2.8. *For the language $\mathcal{L}_\mathcal{R}$, there are two trivial $\mathcal{R}$-term ideals; $\mathbb{T}_{all}$ consisting of all terms of the language and $\mathbb{T}_{cl}$ consisting of all closed terms, with majorizing sets as the set of all polynomials and the whole set of closed terms, respectively. To have a non-trivial example, consider the language of bounded arithmetic extended with subtraction and division and the theory $\mathcal{A}_p$ as* BASIC+$\mathcal{R}$ *plus the axioms $|x| \leq x$, $|xy| \leq |x| + |y|$ and $x \leq y \rightarrow |x| \leq |y|$. Now define $\mathbb{T}_p$ as the class of all terms majorized by a term in the form $p(|\vec{x}|)$ for some polynomial p provably in $\mathcal{A}_p$. The majorizing subset is the set of all terms in the form $p(|\vec{x}|)$ and the reason that the set is an $\mathcal{A}_p$-ideal is that all terms are bounded by a polynomial in length and the fact that these terms are increasing, both provably in $\mathcal{A}_p$.*

Using these ingredients, we can introduce the general definition of a bounded theory of arithmetic:

Definition 2.2.9. *Let $\mathcal{A} \supseteq \mathcal{R}$ be a set of quantifier-free axioms, $\mathbb{T}$ be an $\mathcal{A}$-term ideal and Φ be a class of bounded formulas closed under substitution and subformulas. By the first order bounded arithmetic, $\mathfrak{B}(\mathbb{T}, \Phi, \mathcal{A})$ we mean the theory in the language $\mathcal{L}$ which consists of axioms $\mathcal{A}$, and the $(\mathbb{T}, \Phi)$-induction axiom, i.e.,*

$$A(0) \wedge \forall x(A(x) \rightarrow A(x+1)) \rightarrow \forall x A(t(x)),$$

where $A \in \Phi$ and $t \in \mathbb{T}$. In case that the $\mathcal{A}$-term-ideal $\mathbb{T}$ equals to the set of all terms of the language, we denote the theory $\mathfrak{B}(\mathbb{T}, \Phi, \mathcal{A})$ by $\mathfrak{B}(\Phi, \mathcal{A})$.

Example 2.2.10. *With our definition of bounded arithmetic, different kinds of theories can be considered as bounded theories of arithmetic, for instance $I\Delta_0$, S_n^k, T_n^k, $I\Delta_0(\exp)$ and* PRA *augmented with subtraction and division in the language and the axioms of $\mathcal{R}$ in the theory, are just some of the well-known examples.*

Remark 2.2.11. *Note that the theory $\mathfrak{B}(\mathbb{T}, \Phi, \mathcal{A})$ may not have access to the full-induction scheme*

$$A(0) \wedge \forall x(A(x) \to A(x+1)) \to \forall x A(x),$$

for any $A \in \Phi$. For instance, in the theory S_2^k, the system only has the length-induction that is believed to be weaker than the usual induction in T_2^k.

As usual in the proof theoretical investigations, we are interested in a more structural representation of proofs. For this purpose and for any arbitrary set $\mathbf{Ax}$ of sequents, consider the system $G1(\mathbf{Ax})$ consisting of the following rules:

Axioms:

$$\frac{}{A \Rightarrow A} \qquad \frac{}{A_1, \dots A_n \Rightarrow B_1, \dots B_m}$$

where the right axiom is a substitution of a sequent in $\mathbf{Ax}$.

Structural Rules:

$$(wL)\ \frac{\Gamma \Rightarrow \Delta}{\Gamma, A \Rightarrow \Delta} \qquad (wR)\ \frac{\Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, A}$$

$$(cL)\ \frac{\Gamma, A, A \Rightarrow \Delta}{\Gamma, A \Rightarrow \Delta} \qquad (cR)\ \frac{\Gamma \Rightarrow \Delta, A, A}{\Gamma \Rightarrow \Delta, A}$$

$$(cut)\ \frac{\Gamma_0 \Rightarrow \Delta_0, A \qquad \Gamma_1, A \Rightarrow \Delta_1}{\Gamma_0, \Gamma_1 \Rightarrow \Delta_0, \Delta_1}$$

Propositional Rules:

$$\vee L\ \frac{\Gamma_0, A \Rightarrow \Delta_0 \qquad \Gamma_1, B \Rightarrow \Delta_1}{\Gamma_0, \Gamma_1, A \vee B \Rightarrow \Delta_0, \Delta_1} \qquad \vee R\ \frac{\Gamma \Rightarrow \Delta, A_i}{\Gamma \Rightarrow \Delta, A_0 \vee A_1}\ (i = 0, 1)$$

$$\wedge L\ \frac{\Gamma, A_i \Rightarrow \Delta}{\Gamma, A_0 \wedge A_1 \Rightarrow \Delta}\ (i = 0, 1) \qquad \wedge R\ \frac{\Gamma_0 \Rightarrow \Delta_0, A \qquad \Gamma_1 \Rightarrow \Delta_1, B}{\Gamma_0, \Gamma_1 \Rightarrow \Delta_0, \Delta_1, A \wedge B}$$

$$\to L\ \frac{\Gamma_0 \Rightarrow A, \Delta_0 \qquad \Gamma_1, B \Rightarrow \Delta_1}{\Gamma_0, \Gamma_1, A \to B \Rightarrow \Delta_0, \Delta_1} \qquad \to R\ \frac{\Gamma, A \Rightarrow B, \Delta}{\Gamma \Rightarrow \Delta, A \to B}$$

$$\neg L\ \frac{\Gamma \Rightarrow \Delta, A}{\Gamma, \neg A \Rightarrow \Delta} \qquad \neg R\ \frac{\Gamma, A \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg A}$$

Quantifier rules:

$$\forall L\ \frac{\Gamma, A(s) \Rightarrow \Delta}{\Gamma, \forall y\ A(y) \Rightarrow \Delta} \qquad \forall R\ \frac{\Gamma \Rightarrow \Delta, A(b)}{\Gamma \Rightarrow \Delta, \forall y\ A(y)}$$

$$\exists L\ \frac{\Gamma, A(b) \Rightarrow \Delta}{\Gamma, \exists y\ A(y) \Rightarrow \Delta} \qquad \exists R\ \frac{\Gamma \Rightarrow \Delta, A(s)}{\Gamma, \Rightarrow \Delta, \exists y\ A(y)}$$

Bounded Quantifier rules:

$$\forall^{\leq} L\ \frac{\Gamma, A(s) \Rightarrow \Delta}{\Gamma, s \leq t, \forall y \leq t\ A(y) \Rightarrow \Delta} \qquad \forall^{\leq} R\ \frac{\Gamma, b \leq t \Rightarrow \Delta, A(b)}{\Gamma \Rightarrow \Delta, \forall y \leq t\ A(y)}$$

$$\exists^{\leq} L\ \frac{\Gamma, b \leq t, A(b) \Rightarrow \Delta}{\Gamma, \exists y \leq t\ A(y) \Rightarrow \Delta} \qquad \exists^{\leq} R\ \frac{\Gamma \Rightarrow \Delta, A(s)}{\Gamma, s \leq t, \Rightarrow \Delta, \exists y \leq t\ A(y)}$$

Note that in the rules $(\forall R)$, $(\exists L)$, $(\forall^{\leq} R)$ and $(\exists^{\leq} L)$, the variable b must not occur in the lower sequent of the rule.

There is also another type of sequent calculus, called $G3(\mathbf{Ax})$, absorbing all the structural rules. It is defined with the same rules, by eliminating structural rules and replacing the axioms, the cut rule, the propositional rules and the rules $(\forall L)$, $(\exists R)$, $(\forall^{\leq} L)$ and $(\exists^{\leq} R)$ by the following rules:

Axioms:

$$\frac{}{\Gamma, P \Rightarrow P, \Delta} \qquad \frac{}{\Gamma, P_1, \ldots P_n \Rightarrow Q_1, \ldots Q_m, \Delta}$$

where $P_1, \ldots P_n \Rightarrow Q_1, \ldots Q_m \in cl(\mathbf{Ax})$ and P, P_i's and Q_j's are all atomic formulas. By $cl(\mathbf{Ax})$ we mean the closure of $\mathbf{Ax}$ under substitution and contraction.

Structural Rules:

$$_{(cut)}\frac{\Gamma \Rightarrow \Delta, A \qquad \Gamma, A \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}$$

where P is an atmoic formula, and

Propositional Rules:

$$_{\vee L}\frac{\Gamma, A \Rightarrow \Delta \qquad \Gamma, B \Rightarrow \Delta}{\Gamma, A \vee B \Rightarrow \Delta} \qquad _{\vee R}\frac{\Gamma \Rightarrow \Delta, A, B}{\Gamma \Rightarrow \Delta, A \vee B}$$

$$_{\wedge L}\frac{\Gamma, A, B \Rightarrow \Delta}{\Gamma, A \wedge B \Rightarrow \Delta, C} \qquad _{\wedge R}\frac{\Gamma \Rightarrow \Delta, A \qquad \Gamma \Rightarrow \Delta, B}{\Gamma \Rightarrow \Delta, A \wedge B}$$

$$_{\rightarrow L}\frac{\Gamma \Rightarrow A, \Delta \qquad \Gamma, B \Rightarrow \Delta}{\Gamma, A \rightarrow B \Rightarrow \Delta} \qquad _{\rightarrow R}\frac{\Gamma, A \Rightarrow B, \Delta}{\Gamma \Rightarrow \Delta, A \rightarrow B}$$

Quantifier rules:

$$_{\forall L}\frac{\Gamma, A(s), \forall y\ A(y) \Rightarrow \Delta}{\Gamma, \forall y\ A(y) \Rightarrow \Delta} \qquad _{\exists R}\frac{\Gamma \Rightarrow \Delta, A(s), \exists y\ A(y)}{\Gamma, \Rightarrow \Delta, \exists y\ A(y)}$$

Bounded Quantifier rules:

$$_{\forall^{\leq} L}\frac{\Gamma, A(s), \forall y \leq t\ A(y) \Rightarrow \Delta}{\Gamma, s \leq t, \forall y \leq t\ A(y) \Rightarrow \Delta} \qquad _{\exists^{\leq} R}\frac{\Gamma \Rightarrow \Delta, A(s), \exists y \leq t\ A(y)}{\Gamma, s \leq t, \Rightarrow \Delta, \exists y \leq t\ A(y)}$$

Using the system $G1$, choosing $\mathbf{Ax}$ as the set of all sequents $(\Rightarrow A)$ where $A \in \mathcal{A}$ and adding the following induction rule to $G1(\mathbf{Ax})$:

Induction:

$$_{(Ind)}\frac{\Gamma, A(b) \Rightarrow \Delta, A(b+1)}{\Gamma, A(0) \Rightarrow \Delta, A(t)}$$

for every $A \in \Phi$ and $t \in \mathbb{T}$, we can capture the theory $\mathfrak{B}(\mathbb{T}, \Phi, \mathcal{A})$. Note that in the induction rule the variable b must not occur in the lower sequent of the rule.

The most important property of the sequent calculi that we have defined is cut elimination:

Theorem 2.2.12. *(Cut Elimination)*

(*i*) *Any proof in the systems* $G1(\boldsymbol{Ax})$ *and* $G3(\boldsymbol{Ax})$ *can be transformed to a proof in which every cut rule has at least one premise chosen from the axioms of* $\boldsymbol{Ax}$.

(*ii*) *If* $\mathfrak{B}(\mathbb{T}, \Phi, \mathcal{A}) \vdash \Gamma \Rightarrow \Delta$ *then there exists a free-cut free proof for the same sequent in the same system.*

Corollary 2.2.13. *If* $\Gamma \cup \Delta \subseteq \Phi$ *and* $\mathfrak{B}(\mathbb{T}, \Phi, \mathcal{A}) \vdash \Gamma \Rightarrow \Delta$ *then there exists a proof of the same sequent in the same system such that all formulas occurring in the proof are in* Φ.

The proofs of the Theorem 2.2.12 and the Corollary 2.2.13 can be essentially found in [18] and [8].

2.3 Non-deterministic Flows

In this section, we will develop a computational extracting method, designed specifically for bounded theories of arithmetic. This method is based on two simple ingredients: First reducing any proof in any bounded theory of arithmetic to a single uniform sequence of implications such that these implications become provable in a very weak theory (usually universal induction-free system powerful enough to handle the four basic mathematical operations, addition, subtraction, multiplication and division). And secondly, developing a program interpretation of the Herbrand theorem adopted for our bounded arithmetical language to witness any step in the implications by the terms of the language. These programs that we call *reduction programs* generalize the usual reduction between k-turn games and teacher-student interactive protocols. In fact, they provide a non-deterministic interactive machinery to witness the existential variables by the universal variables using the terms of the language.

As a result of combining these ingredients, we will establish a general method to extract computational information from bounded arithmetical proofs. As an application of this method, we will first propose a characterization of all total search problems of any complexity in any bounded arithmetical theory, especially in the presence of higher smash functions. These theories mimics the higher order bounded theories in a first order setting and hence our characterization can be interpreted as a characterization of all provably total higher search problems in higher order bounded arithmetic. More specifically, we will investigate the bounded statements of the theories S_2^k for $k \geq 1$ to reduce their provability to a polynomially long sequence of reduction programs between k-turn games. We will also apply our technique to reprove the strong witnessing theorem for theories S_2^k. This type of witnessing theorem has been investigated for different bounded

theories (See [19], [6], [4], [13] and [21]). Here, we propose another proof for this result for the hierarchy $\{S_2^k\}_{k\geq 1}$.

2.3.1 Non-deterministic Reductions and Reduction Programs

Let us begin right away by non-deterministic reductions.

Definition 2.3.1. *Let $\mathcal{B}$ be a theory and $A(\vec{x})$ and $B(\vec{x})$ be some formulas in the language $\mathcal{L}$. We say $B(\vec{x})$ is non-deterministically $\mathcal{B}$-reducible to $A(\vec{x})$ and we write $A(\vec{x}) \geq_n^{\mathcal{B}} B(\vec{x})$ if $\mathcal{B} \vdash A(\vec{x}) \to B(\vec{x})$. Moreover, by the equivalence $A \equiv_n^{\mathcal{B}} B$ we mean the conjunction of $A \geq_n^{\mathcal{B}} B$ and $B \geq_n^{\mathcal{B}} A$.*

The natural question is that how this proof-theoretic concept can be considered as a computational reduction and why it is called non-deterministic. To answer this question, first recall that by the flow machinery, we intend to transform any arithmetical proof to a sequence of reductions, and the base theory for those reductions preferably is a simple universal and possibly induction-free theory. Therefore, we can use the Herbrand theorem for each step of the reduction to witness the essentially existential quantifiers in $A \to B$ by its universal quantifiers. This is actually what is happening in the deterministic reductions, but here the difference is the use of $\vee$-expansions in the Herbrand proof. Intuitively, these expansions allow us to use some constant many terms to witness one existential quantifier as opposed to just one term in the case of deterministic reductions. Moreover, expansions make some room for interaction in providing the witnessing terms which makes the concrete witnesses extremely complicated. For these reasons, we call these reductions non-deterministic and in the following we try to state a computational interpretation of Herbrand theorem, tailored specifically for our setting here.

Definition 2.3.2. *Let $\mathcal{L}$ be a language extending the language of $\mathcal{L}_{\mathcal{R}}$. A formula $A(\vec{x})$ is in the prenex bounded form if there exists a quantifier-free formula G_A, such that*

$$A = \forall y_1 \leq p_1(\vec{x}) \exists z_1 \leq q_1(\vec{x}) \forall y_2 \leq p_2(\vec{x}) \dots G_A(\vec{x}, y_1, z_1, y_2, z_2, \dots)$$

Note that all bounding terms depend only on $\vec{x}$ where $\vec{x}$ is the set of all free variables of $A(\vec{x})$. Moreover, we say that the formula is in the k-prenex bounded form when the number of bounded quantifiers is at most k.

Definition 2.3.3. *Let $\mathcal{L}$ be a language extending the language of $\mathcal{L}_{\mathcal{R}}$, the formulas*

$$\{\forall y_{i1} \leq p_{i1}(\vec{x}) \exists z_{i1} \leq q_{i1}(\vec{x}) \forall y_{i2} \leq p_{i2}(\vec{x}) \dots G_i(\vec{x}, y_{i1}, z_{i1}, y_{i2}, z_{i2}, \dots)\}_{i \in I}$$

and

$$\{\forall u_{j1} \leq p'_{j1}(\vec{x}) \exists v_{j1} \leq q'_{j1}(\vec{x}) \forall u_{j2} \leq p'_{j2}(\vec{x}) \dots H_j(\vec{x}, u_{j1}, v_{j1}, u_{j2}, v_{j2}, \dots)\}_{j \in J}$$

be in the prenex bounded form where $\vec{x}$, y_{in}, z_{in}, u_{jm} and v_{jm} are distinct variables and $\mathcal{B}$ be a theory extending $\mathcal{R}$. Define $\mathcal{V}$ as the set of distinct variables y_{in}^k, z_{in}^k, u_{jm}^k and v_{jm}^k for $k \geq 0$. These variables provide infinite many copies of

any variable from y_{in}, z_{in}, u_{jm} and v_{jm}. Moreover, note that for $k = 0$ we have the original copy, i.e., $y^0_{in} = y_{in}$, $z^0_{in} = z_{in}$, $u^0_{jm} = u_{jm}$ and $v^0_{jm} = v_{jm}$.

*Consider the instructions [**Read** $X \leq t(\vec{x})$] and [**Compute** Y **by** $s \leq t(\vec{x})$] where t is a term depending only on $\vec{x}$ and variables X and Y are chosen from the set $\mathcal{V}$. By a $\mathcal{B}$-reduction program from $\{H_j\}_{j\in J}$ to $\{G_i\}_{i\in I}$ we mean a sequence $P = (P_r)^l_{r=0}$ of instructions such that:*

(*i*) *The instruction [**Read** $X \leq t(\vec{x})$] applies only on $X = u^k_{jm}$ and $X = z^k_{in}$ variables.*

(*ii*) *The instruction [**Compute** Y **by** $s \leq t(\vec{x})$] applies only on $Y = v^k_{jm}$ and $Y = y^k_{in}$ variables.*

(*iii*) *Any variable can be read or computed at most once.*

(*iv*) *We can read or compute a variable Z if there exists a decreasing path of already read or computed variables starting from Z and ending in one of the variables $\{y_{i1}\}_{i\in I}$ or $\{u_{j1}\}_{j\in J}$. By "decreasing", we refer to the order defined by the relations $(y^k_{in} \prec z^k_{in})$, $(z^k_{in} \prec y^k_{i(n+1)})$, $(u^k_{jm} \prec v^k_{jm})$, $(v^k_{jm} \prec u^k_{j(m+1)})$ and $(Y^k \prec Y^{k+1})$ for any $Y \in \{v_{jm}, y_{in}\}$.*

(*v*) *In the instruction [**Compute** Y **by** $s \leq t(\vec{x})$], the term s depends only on the variables $\vec{x}$ and the variables that had been read before the current stage. Moreover, we have to have $\mathcal{B} \vdash \forall \vec{X} \leq \vec{r}(\vec{x}) s(\vec{X}) \leq t(\vec{x})$ where $\vec{X}$ are the previously read variables and $\vec{r}(\vec{x})$ are their corresponding bounds.*

(*vi*) *For the last condition, first define $S(P^{<r})$ recursively as:*

$$S(P^{<0}) = \{G_i\}_{i\in I} \Rightarrow \{H_j\}_{j\in J}$$

and $S(P^{<r+1})$ is defined from $S(P^{<r})$ by the following rule:

*There are two cases to consider. First if P_r is the instruction [**Read** $X \leq t(\vec{x})$], then replace all instances of $\forall X \leq t(\vec{x})A(X)$ in the succedent of $S(P^{<r})$ by $A(X)$. And also replace all instances of $\exists X \leq t(\vec{x})A(X)$ in the precedent of $S(P^{<r})$ again by $A(X)$. Second, if P_r is the instruction [**Compute** Y **by** $s \leq t(\vec{x})$], then for any occurrence of $\exists Y \leq t(\vec{x})A(Y)$ in the succedent of $S(P^{<r})$, replace $\exists Y \leq t(\vec{x})A(Y)$ by $(\exists Y \leq t(\vec{x})A(Y))^{+1}$ and add $A(s)$ to the succedent of $S(P^{<r})$. And for any occurrence of $\forall Y \leq t(\vec{x})A(Y)$ in the precedent, replace $\forall Y \leq t(\vec{x})A(Y)$ by $(\forall Y \leq t(\vec{x})A(Y))^{+1}$ and add $A(s)$ in the precedent, where C^{+1} means increasing the upper index of any bounded variable in C by one.*

Now after defining $S(P^{<r})$, we also have to have the following last condition: There should be a quantifier-free sub-sequent $S' = (\Gamma \Rightarrow \Delta)$ of $S(P^{<l+1})$ such that $\forall \vec{X} \leq \vec{r}(\vec{x})(\bigwedge \Gamma \to \bigvee \Delta)$ is provable in $\mathcal{B}$, where $\vec{X}$ are all the read variables occurred in S' and $\vec{r}(\vec{x})$ are their corresponding bounds.

Remark 2.3.4. *(Game Interpretation) Let*

$$C = \forall y_1 \leq p_1(\vec{x}) \exists z_1 \leq q_1(\vec{x}) \forall y_2 \leq p_2(\vec{x}) \dots G_C(\vec{x}, y_1, z_1, y_2, z_2, \dots)$$

be in the k-prenex bounded form with exactly k quantifiers. The game associated to this formula, $\mathcal{G}_C$, is defined as the following: There are two players. The first player chooses a number $y_1 \leq p_1(\vec{x})$, then the second player chooses a number $z_1 \leq q_1(\vec{x})$ and they continue alternately, until they reach the end of the quantifiers. At the end, if $G(\vec{x}, y_1, z_1, y_2, z_2, \dots)$ becomes true the second player wins and otherwise the first player is the winner. Now it is clear that the second player in the game $\mathcal{G}_C$ has a winning strategy iff the formula C is true. With this game interpretation, any reduction program from B to A is nothing but a reduction to transfer a second player's winning strategy in the game $\mathcal{G}_A$ to a winning strategy for him in the game $\mathcal{G}_B$. Note that unlike the usual deterministic reductions between the games, these reduction programs provide a complicated protocol to transfer the winning strategies.

In the following, we will illuminate the notion of a reduction program by some concrete examples.

Example 2.3.5. *(Deterministic Game Reductions) The usual complexity theoretic reduction between k-turn games is a specific example of the reduction programs. To be more precise, assume that $\mathcal{L} = \mathcal{L}_{\mathrm{PV}}$ and $\mathcal{B} = Th(\mathbb{N})$. Now consider*

$$\forall y_1 \leq p_1(\vec{x}) \exists z_1 \leq q_1(\vec{x}) \forall y_2 \leq p_2(\vec{x}) \dots G(\vec{x}, y_1, z_1, y_2, z_2, \dots)$$

and

$$\forall u_1 \leq p'_1(\vec{x}) \exists v_1 \leq q'_1(\vec{x}) \forall u_2 \leq p'_2(\vec{x}) \dots H(\vec{x}, u_1, v_1, u_2, v_2, \dots)$$

with k-many bounded quantifiers and define the following natural PV*-reduction program:*

*[**Read** $u_1 \leq p'_1(\vec{x})$]; [**Compute** y_1 **by** $f_1(\vec{x}, u_1) \leq p_1(\vec{x})$]; [**Read** $z_1 \leq q_1(\vec{x})$]; [**Compute** v_1 **by** $g_1(\vec{x}, u_1, z_1) \leq q'_1(\vec{x})$]; ...*

where f_i's and g_j's are polynomial time computable functions represented by terms in the language $\mathcal{L}_{\mathrm{PV}}$. These reductions that we call deterministic reductions between k-turn games are the simplest example of reduction programs.

Example 2.3.6. *(Non-determinism) Let $\mathcal{L}$ be the language of* PV *and $A(x, y, z)$ be an atomic formula in this language. Now consider the formulas*

$$\exists u \leq s(x) \forall v \leq t(x)\ A(x, u, v)$$

and

$$\exists y y' \leq s(x) \forall z z' \leq t(x)\ [A(x, y, z) \vee A(x, y', z')]$$

in this language. Since these formulas are logically equivalent, it seems quite reasonable to assume that the first formula is reducible to the second one. Moreover, this equivalence is quite elementary and it is just on the level of pure first order logic. Hence, we can expect a very low complexity reduction in this case. Let us try to construct such a possible reduction. (Note that the notion of a reduction

program is defined for formulas in prenex bounded form and since the formula $[A(x,y,z) \vee A(x,y',z')]$ *is not atomic, speaking of reduction programs in this case is not technically correct. However, this is not a serious issue since we can represent the formula* $A(x,y,z)$ *by* $\alpha(x,y,z)=0$ *for some term* α *and hence the formula* $[A(x,y,z) \vee A(x,y',z')]$ *can be safely replaced by* $\alpha(x,y,z)\cdot\alpha(x,y',z')=0$. *Having all said, we still prefer keeping the original non-atomic form to be more explanatory in our discussion on the non-deterministic nature of reductions.) To construct a reduction, we have to take a look at a proof of*

$$\exists u \leq s(x) \forall v \leq t(x)\ A(x,u,v)$$

from

$$\exists y y' \leq s(x) \forall z z' \leq t(x)\ [A(x,y,z) \vee A(x,y',z')].$$

The simplest proof works as the following: Assume we have $y \leq s(x)$ *and* $y' \leq s(x)$ *such that*

$$\forall z z' \leq t(x)[A(x,y,z) \vee A(x,y',z')]$$

which implies

$$\forall z \leq t(x) A(x,y,z) \vee \forall z' \leq t(x) A(x,y',z')$$

Then there are two possibilities: If $\forall z \leq t(x)A(x,y,z)$ *then pick* $u=y$ *and if* $\neg\forall z \leq t(x)A(x,y,z)$ *which implies* $\forall z' \leq t(x)A(x,y',z')$, *pick* $u=y'$.

Simulating this argument by the usual reductions between 3*-turn games, as in Example 2.3.5, (assume the existence of a dummy bounded universal quantifier in the leftmost part of the formulas), we observe that our computational power has to be strong enough to decide* $\forall z \leq t(x)A(x,y,z)$ *to provide such a witness. But since* $\forall z \leq t(x)A(x,y,z)$ *can be extremely complex,* CoNP*-complete for instance, this task is far beyond the usual low complexity power that we can afford. Hence, it seems that finding a deterministic reduction is not that easy, if not impossible.*

Now let us relax the strict structure of the deterministic game reductions to the following weaker non-determinism appeared in the reduction programs: In the process of witnessing, allow reductions to provide possibly more than one candidate and expect at least one of them works at a time. For instance, in this example, provide two different guesses for u *like* $g(x,y,y')=y$ *and* $h(x,y,y')=y'$ *and expect the sequent*

$$\{\forall v^0 \leq t(x)A(x,g(x,y,y'),v^0), \forall v^1 \leq t(x)A(x,h(x,y,y'),v^1)\}$$

to be reducible to

$$\forall z \leq t(x) \forall z' \leq t(x)\ [A(x,y,z) \vee A(x,y',z')]$$

via a PV*-reduction program. For the latter, it is just enough to use the universal quantifiers to witness themselves via identity terms. More formally:*

*[**Read** $y \leq s(x)$]; [**Read** $y' \leq s(x)$]; [**Compute** u^0 **by** $g(x,y,y') \leq s(x)$]; [**Compute** u^1 **by** $h(x,y,y') \leq s(x)$]; [**Read** $v^0 \leq t(x)$]; [**Read** $v^1 \leq t(x)$]; [**Compute** z **by** $v^0 \leq t(x)$]; [**Compute** z' **by** $v^1 \leq t(x)$].*

Hence, we can observe that non-determinism possibly provides more reductions than what the strict determinism can do. Moreover, note that this type of non-determinism is just the computational incarnation of the contraction rule which makes its use somehow unavoidable.

Example 2.3.7. *(Student-Teacher Game) The real power of the reduction programs lies in the combination of non-determinism and interaction. In the Example 2.3.6, we observed the impact of the non-determinism part. In this example we will explain how this non-determinism leads to some complicated interactions. For this purpose, let us interpret the teacher-student game of the KPT theorem ([14]) as an example of our reduction programs. Assume we have the formula* $\exists y \leq t(x) \forall z \leq s(x) A(x,y,z)$ *where* A *is an atomic formula in the language of* PV*. Then consider the following* PV*-reduction program from* $\exists y \leq t(x) \forall z \leq s(x) A(x,y,z)$ *to* $\top$ *with length* $2l$*:*

*[**Compute*** $y = y^0$ ***by*** $f_0(x) \leq t(x)$*]; [**Read*** $z = z^0 \leq s(x)$*]; [**Compute*** y^1 ***by*** $f_1(x, z^0) \leq t(x)$*]; [**Read*** $z^1 \leq s(x)$*]; [**Compute*** y^2 ***by*** $f_2(x, z^0, z^1) \leq t(x)$*]; ...*

Since it is a reduction program, the following is provable in PV*:*

$$A(x, f_0(x), z^0) \vee A(x, f_1(x, z^0), z^1) \vee \ldots \vee A(x, f_l(x, z^0, \ldots, z^{l-1}))$$

with the bounded universal quantifiers on $z^k \leq s(x)$ *for* $k \leq l-1$*. The point in the interaction between the so-called student and teacher is mimicked by computing* y *as* y^0*, reading* $z = z^0$*, computing* y *again under the name* y^1 *but this time with access to* z^0*, reading* z^1 *and computing* y *again under the name* y^2 *but now with more information about both* z^0 *and* z^1 *and so on. This non-determinism that lets us compute a variable finite many times with different functions and the interaction with the inside universal quantifiers to guess the existential quantifier again is the main power of reduction programs.*

Example 2.3.8. *(Impossibility of Simulation) In this example we want to provide an evidence for what we observed in the Example 2.3.6 to show that it is generally impossible to simulate the non-deterministic reductions and reduction programs by usual deterministic reductions. Assume* $A(x,y,z,t) = (y = 0 \wedge B(x,t)) \vee (y = 1 \wedge \neg B(x,z))$ *where* $B(x,t)$ *is an arbitrary atomic formula in the language of* PV*. We want to show that there is no deterministic* $Th(\mathbb{N})$*-reduction from*

$$\exists u \leq 1 \exists v \leq s \forall w \leq s \; A(x,u,v,w)$$

to

$$\exists yy' \leq 1 \exists tt' \leq s \forall zz' \leq s \; [A(x,y,z,t) \vee A(x,y',z',t')]$$

(Note that again, our formulas are not in the prenex bounded form and hence speaking of reduction programs is not technically correct. However, we can resolve the issue as in the Example 2.3.6.) Assume that there exists such a deterministic $Th(\mathbb{N})$*-reduction. Hence, there is a polytime function* f *such that:*

$$\exists v \leq s \forall w \leq s \; A(x, f(x,y,y'), v, w)$$

is reducible to

$$\exists tt' \leq s \forall zz' \leq s [A(x,y,z,t) \vee A(x,y',z',t')]$$

which means that

$$\exists tt' \leq s \forall zz' \leq s[A(x,y,z,t) \vee A(x,y',z',t')]$$

implies

$$\exists v \leq s \forall w \leq s\ A(x, f(x,y,y'), v, w)$$

in $Th(\mathbb{N})$ *for all* $y, y' \leq 1$. *Pick* $y = 0$ *and* $y' = 1$. *It is easy to see that the left side of the implication is true because either* $\exists t \leq s\ B(x,t)$ *or* $\forall z' \leq s\ \neg B(x,z')$ *is true, hence the right side should be true, as well. But the truth of the right side means*

$$(f(x,0,1) = 0 \wedge \exists v \leq s\ B(x,v)) \vee (f(x,0,1) = 1 \wedge \forall w \leq s\ \neg B(x,w))$$

which means that we have a polytime decision procedure for the NP *predicate* $\exists w \leq s\ B(x,v)$ *which implies* $\mathbf{NP} = \mathbf{P}$.

Remark 2.3.9. *The Example 2.3.8 shows that pure logical deductions are far beyond the power of low level deterministic reductions. In other words, it is possible to prove* B *by* A *just by some elementary methods of logic but it does not mean that* B *can be deterministically reducible to* A. *Let us explain where the problem is. At the first glance, it seems that all logical rules are completely syntactical and amenable to low complexity reductions. It is correct everywhere except for one logical rule: the contraction rule which is more or less responsible for all kinds of computational explosions like the explosion of the lengths of the proofs after the elimination of cuts. Notice that the reason that we have the equivalence in the Example 2.3.6 is this contraction rule and it is easy to see that this rule is the source of non-determinism and hence interactions. Therefore, it seems natural to use non-deterministic reductions to simulate computationally what is going on in the realm of proofs.*

Now it is time to relate the proof theoretical non-deterministic reductions to the computational reduction programs. This is the task of our reinterpretation of the generalized Herbrand theorem for the bounded domain:

Theorem 2.3.10. *Let* $\mathcal{B} \supseteq \mathcal{R}$ *be a universal theory and* $A(\vec{x})$ *and* $B(\vec{x})$ *two formulas in the prenex bounded form. Then* $A(\vec{x}) \geq_n^{\mathcal{B}} B(\vec{x})$ *iff there exists a* $\mathcal{B}$*-reduction program from* $B(\vec{x})$ *to* $A(\vec{x})$.

Proof. The proof is based on the fact that any $\mathcal{B}$-reduction program is nothing but a backward interpretation of a proof that consists of some bounded quantifier rules applied on top of a quantifier-free $\mathcal{B}$-provable statement. This backward interpretation transforms the rules $(\forall^{\leq} R)$ and $(\exists^{\leq} L)$ to **Read** instructions and rules $(\exists^{\leq} R)$ and $(\forall^{\leq} L)$ to **Compute** instructions. The rest of this proof is the formalization of this very idea.

1. First assume that there exists a $\mathcal{B}$-reduction program $\{P_r\}_{r=0}^{l}$ from $B(\vec{x})$ to $A(\vec{x})$. We have to prove the following claim:

Claim. We want to show that all the free variables of $S(P^{<k})$ are among $\vec{x}$ or the variables that had been read before k. We prove the claim by induction on k.

For $k = 0$ the claim is clear. For $k + 1$, if P_k is the instruction [**Read** $X \leq t(\vec{x})$], then by definition the free variables of $S(P^{<k+1})$ are among the free variables of $S(P^{<k})$ and X. By IH, all free variables of $S(P^{<k})$ had been read before k and X itself has been read in the stage k, which complete the proof. If P_k is the instruction [**Compute** Y **by** $s \leq t(\vec{x})$], then by definition the free variables of $S(P^{<k+1})$ are among the free variables of $S(P^{<k})$ and the free variables of s. But the free variables of s are among $\vec{x}$ and the variables that had been read before k which is exactly what we wanted to prove.

Now let us come back to prove the theorem. Define $\bar{S}(P^{<k})$ as

$$\forall \vec{X} \leq \vec{r}(\vec{x})[\bigwedge S^p(P^{<k}) \to \bigvee S^s(P^{<k})]$$

where $\vec{X}$ are all the read variables occurred freely in $S(P^{<k})$, $\vec{r}(\vec{x})$ are their corresponding bounds and $S^p(P^{<k})$ and $S^s(P^{<k})$ are the precedent and succedent of $S(P^{<k})$, respectively. By induction on k we will show that $\mathcal{B} \vdash \bar{S}(P^{<l+1-k})$. For $k = 0$ we have the claim from the definition of a program. To prove the claim for $k + 1$, we have two possibilities: First if P_{l-k} is the instruction [**Read** $X \leq t(\vec{x})$], then $S(P^{<l-k+1})$ is defined from $S(P^{<l-k})$ by replacing all instances of $\forall X \leq t(\vec{x})A(X)$ by $A(X)$ in the right hand-side or $\exists X \leq t(\vec{x})A(X)$ by $A(X)$ in the left hand-side. Since any quantifier appears at most once, this formula is unique. On the other hand, since X is read in the stage $l - k$, then by the claim, $S(P^{<l-k})$ does not have a free variable X. By IH, $\mathcal{B} \vdash \bar{S}(P^{<l-k+1})$. Hence, we can introduce the universal bounded quantifier to have $\mathcal{B} \vdash \bar{S}(P^{<l-k})$. If P_{l-k} is the instruction [**Compute** Y **by** $s \leq t(\vec{x})$], then $S(P^{<l-k+1})$ is defined from $S(P^{<l-k})$ by adding $A(s)$ to its right hand-side if there is $\exists Y \leq t(\vec{x})A(Y)$ also in the right hand-side of $S(P^{<l-k})$ or by adding $A(s)$ in the left hand-side of $S(P^{<l-k})$ if $\forall Y \leq t(\vec{x})A(Y)$ is also appeared in its left hand-side. By IH, $\mathcal{B} \vdash \bar{S}(P^{<l-k+1})$. Since $\mathcal{B} \vdash s \leq t(\vec{x})$ we have $\mathcal{B} \vdash \bar{S}(P^{<l-k})$ by the introduction of bounded existential quantifier rules.

Now by induction we can conclude $\mathcal{B} \vdash S(P^{<0}) = [A(\vec{x}) \Rightarrow B(\vec{x})]$ which is what we wanted to prove.

2. For the other direction of the theorem, first note that $\mathcal{B}$ is a universal theory. Therefore, it is possible to develop a $G3$-style calculus for it, by some axioms like $P_1, P_2, \ldots, P_n \Rightarrow Q_1, Q_2, \ldots, Q_m$ where P_i's and Q_j's are atomic formulas. By Theorem 2.2.12 and since $\mathcal{B} \vdash A \to B$ there is a proof for the sequent $A \Rightarrow B$ in which one of the premises of any cut is an axiom. Therefore, since A and B are two formulas in the prenex bounded form, all the rules in the proof will be $G3$-style bounded quantifier rules, axioms and cuts with axioms as one of their premises. Now change the name of the variables in a way that any variable can be occurred in a quantifier at most once, and for this purpose use only the bounded variables of $A \to B$ with their possibly different variants with upper indices. More precisely, it is enough to modify the rules in the proof such that the usual $(\exists^{\leq} R)$ and $(\forall^{\leq} L)$ rules change to the following rules:

$$\frac{\Gamma, A(s), \forall y^{+1} \leq t(\vec{x})\ A^{+1}(y^{+1}) \Rightarrow \Delta}{\Gamma, s \leq t(\vec{x}), \forall y \leq t(\vec{x})\ A(y) \Rightarrow \Delta} \qquad \frac{\Gamma \Rightarrow \Delta, A(s), \exists y^{+1} \leq t(\vec{x})\ A^{+1}(y^{+1})}{\Gamma, s \leq t(\vec{x}), \Rightarrow \Delta, \exists y \leq t(\vec{x})\ A(y)}$$

where C^{+1} means increasing the upper index of any bounded variable in C by one and y^{+1} means increasing the upper index of y by one. Then since any variable occurs at most in one quantifier, we can change all the rules $(\exists^{\leq}L)$ and $(\forall^{\leq}R)$ to:

$$\frac{\Gamma, y \leq t \Rightarrow \Delta, A(y)}{\Gamma \Rightarrow \Delta, \forall y \leq t\ A(y)} \qquad \frac{\Gamma, y \leq t, A(y) \Rightarrow \Delta}{\Gamma, \exists y \leq t\ A(y) \Rightarrow \Delta}$$

in a way that all the sequents in the proof remain $\mathcal{B}$-provable. The main point is that if any variable occurs at most in one quantifier, substituting the eigenvariable b in the rules $(\exists^{\leq}L)$ and $(\forall^{\leq}R)$ by the bounding variable y itself, does not affect the validity of the proof.

Now, by induction on the length of the proof, we will show that if $\Gamma \Rightarrow \Delta$ appears in a stage of this proof, then there is a $\mathcal{B}$-reduction-program $P = \{P_r\}_{r=0}^{l}$ with $S(P^{<0}) = (\Gamma \Rightarrow \Delta)$ using exactly the variables in the proof with the condition that the variable z^m becomes a variant of the variable z^n when $m > n$. (Note that this condition is inconsistent with our naming condition in the definition of the reduction programs which states that the variable z^k should be considered as a variant of z^0 when $k > 0$ and z^0 occurs as a bounded variable in $S(P^{<0})$. However, this is just a change in the names of the variables that makes everything simpler. Therefore, the rest of this proof should be read, up to this change in the naming condition.)

The claim for the axioms is straightforward. For the bounded existential rule, assume that $\Gamma, s \leq t(\vec{x}) \Rightarrow \Delta, \exists y \leq t(\vec{x})A(y)$ is a consequence of $\Gamma \Rightarrow \Delta, A(s), \exists y^{+1} \leq t(\vec{x})A^{+1}(y)$. Then by IH, there exists a program P with the condition that $S(P^{<0}) = (\Gamma \Rightarrow \Delta, A(s), \exists y^{+1} \leq t(\vec{x})A^{+1}(y))$. Define s' as:

$$s' = \begin{cases} s & \text{if } s \leq t(\vec{x}) \\ t(\vec{x}) & \text{if } s > t(\vec{x}) \end{cases}$$

It is possible to find such a term because the language is powerful enough to have the characteristic function for the order predicate as observed in Remark 2.2.2. Moreover, note that $\mathcal{B} \vdash s' \leq t(\vec{x})$. Define $P' = P$ with different initial sequent $S(P'^{<0})$ as $(\Gamma, s \leq t(\vec{x}) \Rightarrow \Delta, A(s'), \exists y^{+1} \leq t(\vec{x})A^{+1}(y))$. The reason that P' is also a reduction program is the following: The sequent $S(P^{<l+1})$ has a quantifier-free $\mathcal{B}$-provable subsequent S'. But the difference between $S(P^{<l+1})$ and $S(P'^{<l+1})$ is in adding the formula $s \leq t(\vec{x})$ in the left hand-side of $S(P^{<l+1})$ and substituting s' for s in A. We know that $\mathcal{B} \vdash s \leq t(\vec{x}) \rightarrow s = s'$. Pick the correspondent of the S' in $S(P'^{<l+1})$ (S' after substitution s' for s) and call it $S'' = \Gamma'' \Rightarrow \Delta''$. Hence, $\mathcal{B} \vdash \Gamma'', s \leq t(\vec{x}) \Rightarrow \Delta''$ which implies that P' is a reduction program.

Now define Q by $Q_r = P'_{r+1}$ for $r \leq l$ and Q_0 as the instruction [**Compute** y **by** $s' \leq t(\vec{x})$] and $S(Q^{<0}) = (\Gamma, s \leq t(\vec{x}) \Rightarrow \Delta, \exists y \leq t(\vec{x})A(y))$. It is pretty clear that Q is a reduction program which proves the claim.

A similar argument also works for the bounded universal quantifier rule. The only case that we have to check is the cut rule. Assume $(\Gamma, \vec{P} \Rightarrow \Delta, \vec{Q})$ is a consequence of $(\Gamma, \vec{P}, R \Rightarrow \Delta, \vec{Q})$ and $(\Gamma, \vec{P} \Rightarrow R, \vec{Q}, \Delta)$ where the first is an instance

of an axiom with the main sequent $(\vec{P}, R \Rightarrow \vec{Q})$ and the second is provable. By IH there exists a program reducing $\{R, \vec{Q}, \Delta\}$ to $\{\Gamma, \vec{P}\}$. This program also essentially works for reducing $\{\vec{Q}, \Delta\}$ to $\{\Gamma, \vec{P}\}$. More precisely, define $P' = P$ with different initial sequent as $S(P'^{<0}) = (\{\Gamma, \vec{P}\} \Rightarrow \{\vec{Q}, \Delta\})$. The only important thing is showing that $S(P'^{<l+1})$ has a quantifier-free $\mathcal{B}$-provable subsequent. From IH we know that there exists a quantifier-free $\mathcal{B}$-provable subsequent of $S(P^{<l+1})$ which we call $S' = (\Gamma' \Rightarrow \Delta')$. Since all $\vec{P}$, $\vec{Q}$ and R are atomic formulas, they will remain intact through the quantifier opening process of the reduction program, hence the difference between $S(P'^{<l+1})$ and $S(P^{<l+1})$ is in one instance of R in the right-hand side of $S(P^{<l+1})$. Moreover, it implies that $(\mathbf{P} \Rightarrow \mathbf{Q})$ is a subsequent of both $S(P^{<l+1})$ and $S(P'^{<l+1})$. Define $S'' = (\Gamma' \Rightarrow \Delta' - \{R\})$. We show that S'' is a $\mathcal{B}$-provable quantifier-free subsequent of $S(P'^{<l+1})$. Since $\vec{P} \subseteq \Gamma'$ and $\vec{Q}, R \subseteq \Delta'$ we have $\mathcal{B} \vdash \Gamma', R \Rightarrow \Delta' - \{R\}$ because it is an instance of the axioms. Since $\mathcal{B} \vdash \Gamma' \Rightarrow \Delta'$ by cut we have $\mathcal{B} \vdash S''$ which completes the proof.

□

2.3.2 Non-deterministic Flows

In the previous subsection, we defined the concept of a reduction which can be considered as a one-step move of the computational content. Now it is time to let it *flow*:

Definition 2.3.11. *Let Π be a π-class, $A(\vec{x}), B(\vec{x}) \in \Pi$, $\mathcal{B} \supseteq \mathcal{R}$ a theory and $\mathbb{T}$ a $\mathcal{B}$-term ideal. A non-deterministic $(\mathbb{T}, \Pi, \mathcal{B})$-flow from $A(\vec{x})$ to $B(\vec{x})$ is a pair (t, H) where $t(\vec{x}) \in \mathbb{T}$ is a term and $H(u, \vec{x}) \in \Pi$ is a formula such that the following statements are provable in $\mathcal{B}$:*

(i) $H(0, \vec{x}) \leftrightarrow A(\vec{x})$.

(ii) $H(t(\vec{x}), \vec{x}) \leftrightarrow B(\vec{x})$.

(iii) $\forall u < t(\vec{x})\ H(u, \vec{x}) \rightarrow H(u+1, \vec{x})$.

If there exists a non-deterministic $(\mathbb{T}, \Pi, \mathcal{B})$-flow from $A(\vec{x})$ to $B(\vec{x})$ we will write $A(\vec{x}) \rhd_n^{(\mathbb{T},\Pi,\mathcal{B})} B(\vec{x})$. Moreover, if Γ and Δ are sequents of formulas in Π, by $\Gamma \rhd_n^{(\mathbb{T},\Pi,\mathcal{B})} \Delta$ we mean $\bigwedge \Gamma \rhd_n^{(\mathbb{T},\Pi,\mathcal{B})} \bigvee \Delta$. The case for $(\mathbb{T}, \Sigma, \mathcal{B})$-flows is defined similarly by changing Π everywhere with Σ.

Convention. In the remaining part of this section, we will fix an arbitrary choice for the type of a flow as $(\mathbb{T}, \Sigma, \mathcal{B})$-flow or $(\mathbb{T}, \Pi, \mathcal{B})$-flow. For simplicity, and to address both cases simultaneously, we will use the letters Φ and ϕ, standing for a fixed choice from two cases [$\Phi = \Sigma$ and $\phi = \sigma$] or [$\Phi = \Pi$ and $\phi = \pi$]. For instance, by the sentence "Φ is a ϕ-class" we mean either "Σ is a σ-class" or "Π is a π-class". Moreover, we use the shorthand $\rhd$ for $\rhd_n^{(\mathbb{T},\Phi,\mathcal{B})}$ for simplicity and if emphasis on some parts of the triple $(\mathbb{T}, \Phi, \mathcal{B})$ becomes needed, we put back those parts as the superscript of $\rhd$. For instance, if we write $\rhd^{\Phi}$, we want to emphasize on the class of the flow.

The following theorem is the main theorem of the theory of non-deterministic flows for bounded theories of arithmetic.

Theorem 2.3.12. *Let Φ be a ϕ-class, $\Gamma(\vec{x})\cup\Delta(\vec{x}) \subseteq \Phi$ and $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathfrak{B}(\mathbb{T}, \Phi, \mathcal{A})$. Then $\mathfrak{B}(\mathbb{T}, \Phi, \mathcal{A}) \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$ iff $\Gamma \rhd_n^{(\mathbb{T},\Phi,\mathcal{B})} \Delta$.*

To prove this theorem we need the following sequence of lemmas. These lemmas provide a high level calculus for the relation $\rhd$ which makes its use more effective in any practical situation.

Lemma 2.3.13. *(i) (Weak Gluing) If $A(\vec{x}) \rhd B(\vec{x})$ and $B(\vec{x}) \rhd C(\vec{x})$ then $A(\vec{x}) \rhd C(\vec{x})$.*

(ii) (Strong Gluing) If $A(y, \vec{x}) \rhd A(y + 1, \vec{x})$ and $s \in \mathbb{T}$, then $A(0, \vec{x}) \rhd A(s, \vec{x})$.

Proof. For (i) since $A(\vec{x}) \rhd_n B(\vec{x})$ there exists a term $t(\vec{x}) \in \mathbb{T}$, a formula $H(u, \vec{x}) \in \Phi$ such that $\mathcal{B}$ proves the conditions in the Definition 2.3.11. On the other hand since $B(\vec{x}) \rhd_n C(\vec{x})$ we have the corresponding data for $B(\vec{x})$ to $C(\vec{x})$ which we show by $t'(\vec{x})$ and $H'(u, \vec{x})$. Define $r(\vec{x}) = t(\vec{x}) + t'(\vec{x}) + 1$ and

$$I(u, \vec{x}) = \begin{cases} H(u, \vec{x}) & \text{if } u \leq t(\vec{x}) \\ B(\vec{x}) & \text{if } u = t(\vec{x}) + 1 \\ H'(u \dot{-} t(\vec{x}) \dot{-} 2, \vec{x}) & \text{if } t(\vec{x}) + 1 < u \leq t(\vec{x}) + t'(\vec{x}) + 1 \end{cases}$$

Then, it is easy to check that this new data is a non-deterministic $(\mathbb{T}, \Phi, \mathcal{B})$-flow from $A(\vec{x})$ to $C(\vec{x})$. Notice that since $\mathbb{T}$ is closed under successor and addition and $t, t' \in \mathbb{T}$, we have $r \in \mathbb{T}$.

For (ii), if we have $A(y, \vec{x}) \rhd_n A(y + 1, \vec{x})$ it is enough to glue all copies of the sequences of reductions for $0 \leq y \leq s$, to have $A(0, \vec{x}) \rhd_n A(s, \vec{x})$. More precisely, assume that all reductions have the same length $t'(\vec{x})$ greater than $t(s, \vec{x})$. This is an immediate consequence of the facts that we can find a monotone majorization for $t(y, \vec{x})$ like $r(y, \vec{x})$, and since $y \leq s$ we have $t(y, \vec{x}) \leq r(y, \vec{x}) \leq r(s, \vec{x})$. Now it is enough to repeat the last formula in the flow to make the flow longer to reach the length $t'(\vec{x}, \vec{z}) = r(s, \vec{x})$ where $\vec{z}$ is a vector of variables in s. Now, define $t''(\vec{x}, \vec{z}) = s \times (t'(\vec{x}) + 2)$,

$$I(u, \vec{x}) = \begin{cases} H(u, y, \vec{x}) & \text{if } y(t' + 2) < u < (y + 1)(t' + 2) \\ A(y, \vec{x}) & \text{if } u = y(t' + 2) \end{cases}$$

and

$$F(u) = \begin{cases} F(u, y) & \text{if } y(t' + 2) < u < (y + 1)(t' + 2) \dot{-} 1 \\ E_0(u, y) & \text{if } u = y(t' + 2) \\ G_1(u, y + 1) & \text{if } u = (y + 1)(t' + 2) \dot{-} 1 \end{cases}$$

It is easy to see that this new sequence is a non-deterministic $(\mathbb{T}, \Phi, \mathcal{B})$-flow from $A(0, \vec{x})$ to $A(s, \vec{x})$. Notice that $\mathbb{T}$ is closed under substitution, sum and product and therefore, $t'' \in \mathbb{T}$. □

Lemma 2.3.14. *(Conjunction and Disjunction Rules)*

(i) If $\Gamma, A \rhd \Delta$ or $\Gamma, B \rhd \Delta$ then $\Gamma, A \wedge B \rhd \Delta$.

(ii) If $\Gamma_0 \rhd \Delta_0, A$ and $\Gamma_1 \rhd \Delta_1, B$ then $\Gamma_0, \Gamma_1 \rhd \Delta_0, \Delta_1, A \wedge B$.

(*iii*) *If* $\Gamma \rhd \Delta, A$ *or* $\Gamma \rhd \Delta, B$ *then* $\Gamma \rhd \Delta, A \vee B$.

(*iv*) *If* $\Gamma_0, A \rhd \Delta_0$ *and* $\Gamma_1, B \rhd \Delta_1$ *then* $\Gamma_0, \Gamma_1, A \vee B \rhd \Delta_0, \Delta_1$.

Proof. (*i*) and (*iii*), are trivial simply because firstly we have $A \wedge B \geq A$, $A \wedge B \geq B$, $A \geq A \vee B$ and $B \geq A \vee B$ and then we can add the needed formula in the beginning or the end of the flow.

For (*ii*) and (*iv*), we will prove (*ii*), (*iv*) is just dual to (*ii*). If $\Gamma_0 \rhd \Delta_0, A$, then clearly we have $\bigwedge \Gamma_0 \wedge \bigwedge \Gamma_1 \rhd (\bigvee \Delta_0 \vee A) \wedge \bigwedge \Gamma_1$. Moreover, we have $\bigwedge \Gamma_1 \rhd \bigvee \Delta_1 \vee B$ and again we have $\bigwedge \Gamma_1 \wedge (\bigvee \Delta_0 \vee A) \rhd (\bigvee \Delta_1 \vee B) \wedge (\bigvee \Delta_0 \vee A)$. Therefore by weak gluing

$$\bigwedge \Gamma_0 \wedge \bigwedge \Gamma_1 \rhd (\bigvee \Delta_1 \vee B) \wedge (\bigvee \Delta_0 \vee A).$$

But it is easy to see that

$$(\bigvee \Delta_1 \vee B) \wedge (\bigvee \Delta_0 \vee A) \geq_n \bigvee \Delta_1 \vee \bigvee \Delta_0 \vee (A \wedge B).$$

Hence

$$\Gamma_0, \Gamma_1 \rhd \Delta_0, \Delta_1, (A \wedge B).$$

□

In the following, wherever we write $\neg A$, we mean any possible formula B such that $\vdash \neg A \leftrightarrow B$.

Lemma 2.3.15. *(Negation Rules) If* $\Gamma, \Delta \subseteq \Phi$ *and* $A, \neg A \in \Phi$ *then*

(*i*) *If* $\Gamma, A \rhd^\Phi \Delta$ *then* $\Gamma \rhd^\Phi \Delta, \neg A$.

(*ii*) *If* $\Gamma \rhd^\Phi \Delta, A$ *then* $\Gamma, \neg A \rhd^\Phi \Delta$.

Proof. We will prove (*i*), (*ii*) is similar. Since $\Gamma, A \rhd^\Phi \Delta$ there exists $t \in \mathbb{T}$ and $H \in \Phi$ such that the conditions of the Definition 2.3.11 hold. Now, use $H \wedge \neg A$ as the formula to have a flow from $(\bigwedge \Gamma \wedge A) \vee \neg A$ to $\bigvee \Delta \vee \neg A$. Since

$$\mathcal{B} \vdash \bigwedge \Gamma \to (\bigwedge \Gamma \wedge A) \vee \neg A$$

by adding $\bigwedge \Gamma$ to the beginning of the flow we have a flow from Γ to $\Delta, \neg A$. □

Remark 2.3.16. *Note that the cut and induction rules are derivable in the presence of the structural and propositional rules and their context-free versions, i.e,*

$$\frac{A \Rightarrow B \qquad B \Rightarrow C}{A \Rightarrow C} \qquad \frac{A(y) \Rightarrow A(y+1)}{A(0) \Rightarrow A(t)}$$

Therefore since we have weak and strong gluing lemmas, we do not need to prove cut and induction in a separate lemma.

Lemma 2.3.17. *(Implication Rules) If* $A \to B \in \Phi$*:*

(*i*) *If* $\Gamma_0 \rhd^\Phi \Delta_0, A$ *and* $\Gamma_1, B \rhd^\Phi \Delta_1$ *then* $\Gamma_0, \Gamma_1, A \to B \rhd^\Phi \Delta_0, \Delta_1$.

(*ii*) *If* $\Gamma, A \rhd^\Phi \Delta, B$ *then* $\Gamma \rhd^\Phi \Delta, A \to B$.

Proof. Note that when $A \to B \in \Pi$ then since Π is closed under subformulas, we have $A, B \in \Pi$. For (i), since $\Gamma_0 \rhd \Delta_0, A$ by applying conjunction with $A \to B$ everywhere in the flow, we have

$$\bigwedge \Gamma_0 \wedge (A \to B) \rhd (\bigvee \Delta_0 \vee A) \wedge (A \to B).$$

Since

$$(\bigvee \Delta_0 \vee A) \wedge (A \to B) \rhd \bigvee \Delta_0 \vee (A \wedge (A \to B)),$$

and $A \wedge A \to B \geq_n B$, we have

$$\bigvee \Delta_0 \vee (A \wedge (A \to B)) \rhd \bigvee \Delta_0 \vee B.$$

And then since $\Gamma_1 \rhd B, \Delta_1$, by cut on B we have

$$\Gamma_0, \Gamma_1, A \to B \rhd \Delta_0, \Delta_1.$$

For (ii), if $\Gamma, A \rhd B, \Delta$, then by applying disjunction with $A \to B$ everywhere in the flow,

$$(\bigwedge \Gamma \wedge A) \vee (A \to B) \rhd \bigvee \Delta \vee B \vee (A \to B).$$

And since

$$((\bigwedge \Gamma \vee (A \to B)) \wedge (A \vee (A \to B)) \rhd (\bigwedge \Gamma \wedge A) \vee (A \to B),$$

we have

$$((\bigwedge \Gamma \vee (A \to B)) \wedge (A \vee (A \to B)) \rhd \bigvee \Delta \vee B \vee (A \to B).$$

Since $B \geq_n (A \to B)$, by contraction and cut we have $B \vee (A \to B) \rhd A \to B$. On the other hand, $\geq A \vee (A \to B)$. Hence

$$\Gamma \rhd ((\bigwedge \Gamma \vee (A \to B)) \wedge (A \vee (A \to B)),$$

and therefore by gluing $\Gamma \rhd \Delta, A \to B$. □

Now we are ready to prove the following soundness theorem as the first half of the main theorem:

Theorem 2.3.18. *(Soundness) If Φ is a ϕ-class, $\Gamma(\vec{x}) \cup \Delta(\vec{x}) \subseteq \Phi$, $\mathfrak{B}(\mathbb{T}, \Phi, \mathcal{A}) \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$ and $\mathcal{A} \subseteq \mathcal{B}$ then $\Gamma \rhd_n^{(\mathbb{T}, \Phi, \mathcal{B})} \Delta$.*

Proof. We assume $\Phi = \Pi$ is a π-type class, the other case is similar. To prove the theorem we use induction on the length of the free-cut free proof of $\Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$. The importance of using the free-cut free proof is its usual consequence that all the formulas occurring in the proof belong to the class Π itself. (See Corollary 2.2.13.) Since Π consists of bounded formulas, it also implies that the only used quantifier rules are bounded quantifier rules. Hence, we have the following cases:

1. (Axioms). If $\Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$ is a logical axiom then the claim is trivial. If it is a non-logical axiom then the claim will be also trivial because all non-logical axioms are quantifier-free and provable in $\mathcal{B}$. Therefore there is nothing to prove.

2. (Structural Rules). We will prove the case of the contraction rule, the rest are similar. Assume that $\Gamma, A \Rightarrow \Delta$ is prove by left contraction from $\Gamma, A, A \Rightarrow \Delta$. Then by IH, there exists a flow from $\bigwedge \Gamma \wedge (A \wedge A)$ to $\bigvee \Delta$. Since $\mathcal{B} \vdash A \to A \wedge A$, we know

$$\bigwedge \Gamma \wedge A \geq \bigwedge \Gamma \wedge (A \wedge A)$$

adding $\bigwedge \Gamma \wedge A$ to the beginning of the flow, we will have a flow from $\bigwedge \Gamma \wedge A$ to $\bigvee \Delta$ which proves what we wanted.

3. (Cut). See the Remark 2.3.16.

4. (Propositional Rules). The conjunction and disjunction cases are proved in the Lemma 2.3.14. The implication and negation cases are proved in the Lemmas 2.3.15 and 2.3.17, respectively.

5. (Bounded Universal Quantifier Rules, Right). If

$$\Gamma(\vec{x}) \Rightarrow \Delta(\vec{x}), \forall z \leq p(\vec{x}) B(\vec{x}, z)$$

is proved by the $\forall^{\leq} R$ rule by $\Gamma(\vec{x}), b \leq p(\vec{x}) \Rightarrow \Delta(\vec{x}), B(\vec{x}, b)$, then by IH we have $\Gamma(\vec{x}), b \leq p(\vec{x}) \rhd \Delta(\vec{x}), B(\vec{x}, b)$. Therefore, there exists a term $t(\vec{x}, b) \in \mathbb{T}$, a formula $H(u, \vec{x}, b) \in \Pi$ such that the conditions of the Definition 2.3.11 are provable in $\mathcal{B}$. First of all, extend the sequence by repeating the last formula to reach a majorization $s(\vec{x}, b)$. Then, define $t'(\vec{x}) = s(\vec{x}, p(\vec{x}))$ and $H'(u, \vec{x}) = \forall b \leq p(\vec{x}) I(u, \vec{x}, b)$ where

$$I(u, \vec{x}, b) = \begin{cases} H(u, b, \vec{x}) & u \leq s(\vec{x}, b) \\ H(s(\vec{x}, b), b, \vec{x}) & \text{o.w.} \end{cases}$$

Firstly, it is clear that

$$H'(0, \vec{x}) \equiv_n \forall b \leq p(\vec{x}) [\bigwedge \Gamma(\vec{x}) \wedge b \leq p(\vec{x})]$$

because $0 \leq s(\vec{x}, b)$ and hence $I(0, \vec{x}, b) = H(0, b, \vec{x})$. Secondly, note that we have $H'(u, \vec{x}) \geq_n H'(u + 1, \vec{x})$. Thirdly,

$$H'(t'(\vec{x}), \vec{x}) \equiv_n \forall b \leq p(\vec{x}) [\bigvee \Delta(\vec{x}) \vee B(\vec{x}, b)]$$

The reason is as the following: Assume $b \leq t$ then by the monotonicity of s we have $s(\vec{x}, b) \leq t'(\vec{x})$ which implies $I(t'(\vec{x}), \vec{x}, b) = I(s(\vec{x}, b), b, \vec{x})$ and since $I(s(\vec{x}, b), b, \vec{x}) = H(s(\vec{x}, b), b, \vec{x})$ is $\mathcal{B}$-equivalent to $[\bigvee \Delta(\vec{x}) \vee B(\vec{x}, b)]$, the claim follows. Since $t'(\vec{x})$ is constructed by majorizing and substitution from $t \in \mathbb{T}$, it is also in $\mathbb{T}$. Therefore, (H', t') provides a $(\mathbb{T}, \Pi, \mathcal{B})$-flow from

$$\forall b \leq p(\vec{x}) [\bigwedge \Gamma(\vec{x}) \wedge b \leq p(\vec{x})]$$

to

$$\forall b \leq p(\vec{x}) [\bigvee \Delta(\vec{x}) \vee B(\vec{x}, b)$$

Finally add $\bigwedge \Gamma$ to the beginning of the flow and add $\forall b \leq p(\vec{x}) B(\vec{x}, b) \vee \bigvee \Delta$ to its end, then the new flow would be a flow from Γ to $\Delta(\vec{x}), \forall b \leq p(\vec{x}) B(\vec{x}, b)$.

Since the name of a variable does not affect the nature of a reduction, we can complete this part.

6. (Bounded Universal Quantifier Rules, Left). Suppose

$$\Gamma(\vec{x}), s(\vec{x}) \leq p(\vec{x}), \forall z \leq p(\vec{x}) B(\vec{x}, z) \Rightarrow \Delta(\vec{x})$$

is proved by the $\forall^{\leq} L$ rule by $\Gamma(\vec{x}), B(\vec{x}, s(\vec{x})) \Rightarrow \Delta(\vec{x})$. Since $\mathcal{B} \vdash s(\vec{x}) \leq p(\vec{x}) \wedge \forall z \leq p(\vec{x}) B(\vec{x}, z) \rightarrow B(\vec{x}, s(\vec{x}))$, we have

$$s(\vec{x}) \leq p(\vec{x}), \forall z \leq p(\vec{x}) B(\vec{x}, z) \geq B(\vec{x}, s(\vec{x})).$$

Since

$$\Gamma(\vec{x}), B(\vec{x}, s(\vec{x})) \rhd \Delta(\vec{x}),$$

by cut we have

$$\Gamma(\vec{x}), s(\vec{x}) \leq p(\vec{x}), \forall z \leq p(\vec{x}) B(\vec{x}, s(\vec{x})) \rhd \Delta(\vec{x}).$$

Moreover, note that t'' is constructed by majorizing, substitution and successor from $t \in \mathbb{T}$, hence $t'' \in \mathbb{T}$.

7. (Bounded Existential Quantifier Rules, Right). It is similar to 6.

8. (Bounded Existential Quantifier Rules, Left). If $\Gamma, \exists y \leq p(\vec{x}) B(\vec{x}, y) \Rightarrow \Delta$ is proved by the $\exists^{\leq} L$ rule by $\Gamma, b \leq p(\vec{x}), B(\vec{x}, b) \Rightarrow \Delta$, by IH we have $\Gamma, b \leq p(\vec{x}), B(\vec{x}, b) \rhd \Delta$ then since $\exists y \leq p(\vec{x}) B(\vec{x}, y) \in \Pi$, $B(\vec{x}, y)$ has a negation in Π. Since Π is closed under substitution, $B(\vec{x}, b)$ also has a negation in Π. Therefore, by Lemma 2.3.15

$$\Gamma, b \leq p(\vec{x}) \rhd \Delta, \neg B(\vec{x}, b)$$

by 5, we have

$$\Gamma \rhd \Delta, \forall y \leq p(\vec{x}) \neg B(\vec{x}, y)$$

Finally again by Lemma 2.3.15 we have

$$\Gamma, \exists y \leq p(\vec{x}) B(\vec{x}, y) \rhd \Delta.$$

9. (Induction). See the Remark 2.3.16. □

We also have the following completeness theorem:

Theorem 2.3.19. *(Completeness) If* $\Gamma(\vec{x}) \rhd_n^{(\mathbb{T}, \Phi, \mathcal{B})} \Delta(\vec{x})$ *and* $\mathcal{B} \subseteq \mathfrak{B}(\mathbb{T}, \Phi, \mathcal{A})$*, then* $\mathfrak{B}(\mathbb{T}, \Phi, \mathcal{A}) \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$*.*

Proof. If $\Gamma(\vec{x}) \rhd_n^{(\mathbb{T}, \Pi, \mathcal{B})} \Delta(\vec{x})$, then by Definition 2.3.1, there exist a term $t(\vec{x}) \in \mathbb{T}$, and a formula $H(u, \vec{x}) \in \Pi$ such that we have the following:

(*i*) $\mathcal{B} \vdash H(0, \vec{x}) \leftrightarrow \bigwedge \Gamma(\vec{x})$,

(*ii*) $\mathcal{B} \vdash H(t(x), \vec{x}) \leftrightarrow \bigvee \Delta(\vec{x})$,

(*iii*) $\mathcal{B} \vdash \forall u \leq t(\vec{x})\ H(u, \vec{x}) \rightarrow H(u+1, \vec{x})$.

Since $\mathcal{B} \subseteq \mathfrak{B}(\mathbb{T}, \Pi, \mathcal{A})$, we have

$$\mathfrak{B}(\mathbb{T}, \Pi, \mathcal{A}) \vdash \forall u \leq t(\vec{x})\ H(u, \vec{x}) \rightarrow H(u+1, \vec{x}).$$

Since $H(u, \vec{x}) \in \Pi$ and $t \in \mathbb{T}$, by induction we have ,

$$\mathfrak{B}(\mathbb{T}, \Pi, \mathcal{A}) \vdash H(0, \vec{x}) \rightarrow H(t(\vec{x}), \vec{x}).$$

On the other hand, we have $\mathcal{B} \vdash H(0, \vec{x}) \leftrightarrow \bigwedge \Gamma(\vec{x})$ and $\mathcal{B} \vdash H(t(\vec{x}), \vec{x}) \leftrightarrow \bigvee \Delta(\vec{x})$. Therefore, $\mathfrak{B}(\mathbb{T}, \Pi, \mathcal{A}) \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$. □

2.3.3 Applications

In this subsection we will explain some applications of the theory of non-deterministic flows. For this purpose, let us first define a hierarchy of theories of bounded arithmetic to have a variety of theories with different induction lengths for which the non-determinism is the most effective trick. For this purpose, consider the language $\mathcal{L}_n$ as the Buss' language of bounded arithmetic, [8], augmented with subtraction, division and the function symbols $\{\#_k\}_{k \leq n}$ and define BASIC_n as the theory $\mathcal{A}_p$ as in the Example 2.2.8, together with the defining axioms for these new function symbols. These axioms include the axioms of the theory $\mathcal{R}$ and a suitable representation of $x \#_{k+1} y = 2^{|x| \#_k |y|}$. For $m \leq n-1$, define $\mathbb{T}_{n,m}$ as the set consisting of all terms less than the terms of the form $|t|_m$, provably in BASIC_n where $|t|_m$ means applying the length function m many times. We claim that $\mathbb{T}_{n,m}$ is a BASIC_n-term ideal. First note that for any terms t and s, BASIC_n proves that $|t|_m \cdot |s|_m \leq |t \#_{m+1} s|_m$, hence by $m+1 \leq n$, it is easy to prove that $\mathbb{T}_{n,m}$ is closed under addition, multiplication, subtraction and division. Secondly, it is clear that this set is closed under substitutions simply because of its form and finally note that the set has the majorizing terms of the form $|t|_m$ where t just consists of increasing function symbols, i.e. all the function symbols excluding subtraction and division. Now, define the theory $R^k_{m,n}$ as the bounded arithmetic $\mathfrak{B}(\mathbb{T}_m, \hat{\Pi}^b_k(\#_n), \mathrm{BASIC}_n)$.

In the following theorem, we show that it is possible to decompose proofs of $R^k_{m,n}$:

Theorem 2.3.20. *Let* $\Gamma, \Delta \subseteq \hat{\Pi}^b_k(\#_n)$*, then* $R^k_{m,n} \vdash \Gamma \Rightarrow \Delta$ *iff*

$$\Gamma \rhd_n^{(\mathbb{T}_m, \hat{\Pi}^b_k(\#n), \mathrm{BASIC}_n)} \Delta.$$

Note that more smash functions can simulate higher order objects in our first order setting. For instance the theory $R^k_{n-1,n+r}$ can be read as a theory powerful enough to talk about the n-th order objects, has first order induction for the formulas with k-many alternations of these higher order objects and finally has the $\#_{r+1}$ function on the first order elements. For instance having a characterization of $S^k_3 = R^k_{1,3}$-provable sentences of the form $\forall x \exists y \leq |t(x)|\ A(x,y)$ where A is quantifier-free in the language augmented with all computable functions in time $|t|$ where $t \in \mathcal{L}(\#_3)$, is equivalent to providing a characterization of the total NP-search problems of the second order hierarchy V^k_2.

Using the Theorem 2.3.20 for the usual bounded theories S_2^k, we can provide an example of the combination of the ingredients that we have mentioned in the beginning of this section, i.e., first transforming a proof to a sequence of implications over a universal theory and then using the Theorem 2.3.10 to bring the computational content of each implication.

Corollary 2.3.21. *Let $A(\vec{x}), B(\vec{x}) \in \mathcal{L}_{\mathrm{PV}}$ be two formulas in the k-prenex bounded form and $S_2^k(\mathrm{PV})$ be the theory S_2^k written in the language of* PV*. Then $S_2^k(\mathrm{PV}) \vdash A(\vec{x}) \to B(\vec{x})$ iff there exists a polynomial p, a formula $G(u, \vec{x})$ in the k-prenex bounded form with bounds depending only on $\vec{x}$, a uniform sequence of* PV*-reduction programs P_u from $G(u, \vec{x})$ to $G(u+1, \vec{x})$ and two* PV*-reduction programs, one from $A(\vec{x})$ to $G(0, \vec{x})$ and the other from $G(p(|\vec{x}|), \vec{x})$ to $B(\vec{x})$.*

Proof. Since $S_2^k(\mathrm{PV})$ is axiomatizable by $(\mathbb{T}_{1,2}, \hat{\Pi}_k^b)$-induction; all quantifier-free formulas have PV-equivalent atomic representation and PV is a universal theory, the claim is a clear consequence of Theorem 2.3.20 and Theorem 2.3.10. □

Remark 2.3.22. *Note that this theorem transforms the provability of implications of $\hat{\Pi}_k^b$ formulas (written in their k-prenex bounded forms) in S_2^k (written in the language of* PV*) to the existence of polynomially long sequence of k-turn games with a uniform sequence of* PV*-reduction programs between them. This characterization is more or less similar to the characterizations of [19] and [21] for the theories T_2^k. The difference is on the length of the sequences which in our case is polynomial and hence exponentially shorter that the exponentially long sequence of reductions of [19] and [21]. However, our reduction steps are non-deterministic and hence far more complicated than the simple deterministic reductions of [19] and [21]. Using the $\hat{\Sigma}_{k+1}^b$-conservativity of S_2^{k+1} over T_2^k, we can use both characterizations for both theories for appropriate complexity. This technique pushes the previously known characterization of $\forall\hat{\Sigma}_j^b$ consequences of T_2^k for $1 \leq j \leq k$ ([21]), one level up to provide also a characterization of $\forall\hat{\Sigma}_{k+1}^b$ consequences of T_2^k. It is also worth mentioning that we can apply our characterization to provide another combinatorial characterization of the total* NP *search problems of the theory S_2^{k+1}, and hence of T_2^k, based on polynomially long sequence of* PV*-reduction-programs.*

For the second application, we propose a new proof for the strong version of witnessing theorems for the hierarchy S_2^k. This type of strong witnessing theorems appeared in [19], [6], [4], [13] and [21] for different bounded theories including the theories S_2^k.

Define the hierarchy of function classes $\square_k^p$ as: $\square_1^p = \mathrm{FP}$, $\square_{k+1}^p = \mathrm{FP}^{\Sigma_k^p}$ and let $comp(\vec{x}, M, w)$ be a polytime formalization for "*w is a computation of the algorithm M on the inputs $\vec{x}$*" and $out(w)$ be a polynomial time function symbol which reads w and computes the output of w. Then:

Corollary 2.3.23. *(Strong Witnessing Theorem) The provably $\hat{\Sigma}_k^b$-definable functions of S_2^k are in $\square_k^p$, provably in* PV*, i.e. if $S_2^k \vdash \forall\vec{x}\exists y A(\vec{x}, y)$ where $A(\vec{x}, y) \in \hat{\Sigma}_k^b$, then there exists a machine M computing a function $f \in \square_k^p$ such that $\mathrm{PV} \vdash comp(\vec{x}, M, w) \to A(\vec{x}, out(w))$.*

Proof. Assume $S_2^k \vdash \forall \vec{x} \exists y A(\vec{x}, y)$. By Parikh theorem we know that there exists a bound for the existential quantifier. Hence there exists a term $t(\vec{x})$ such that $S_2^k \vdash \forall y \leq t(\vec{x})\ \neg A(\vec{x}, y) \Rightarrow \bot$. W.l.o.g, assume that the language extends the language of PV. Hence, S_2^k is axiomatizable in this language by $(\mathbb{T}_{1,2}, \hat{\Pi}_k^b)$-induction. By Theorem 2.3.20, there exist a polynomial $p(|\vec{x}|)$ and a formula $H(u, \vec{x}) \in \hat{\Pi}_k^b$ such that the following statements are provable in PV:

(*i*) $H(0, \vec{x}) \leftrightarrow [\forall y \leq t(\vec{x})\ \neg A(\vec{x}, y)]$.

(*ii*) $H(p(|\vec{x}|), \vec{x}) \leftrightarrow \bot$.

(*iii*) $\forall u < p(|\vec{x}|)\ H(u, \vec{x}) \rightarrow H(u+1, \vec{x})$.

W.l.o.g we can assume that H is in the k-prenex bounded form. Hence, $H(u, \vec{x}) = \forall z \leq s(\vec{x})\ G(u, \vec{x}, z)$ where $G(u, \vec{x}, z)$ begins with a bounded existential quantifier and hence is in Σ_{k-1}^b. Since PV is a universal theory, by Theorem 2.3.10, there exist a uniform PV-reduction program P_u from $H(u+1, \vec{x})$ to $H(u, \vec{x})$ for $u < p(|\vec{x}|)$; a PV-reduction program N from $H(0, \vec{x})$ to $\forall y \leq t(x)\ \neg A(x, y)$ and finally a PV-reduction program K from $\bot$ to $H(p(|\vec{x}|), \vec{x})$. The idea is using the power to decide Σ_{k-1}^b formulas and in needed cases finding the witnesses for those decisions, to simplify the reduction programs. We will simplify the reduction program from $H(u+1, \vec{x})$ to $H(u, \vec{x})$, the cases for the other two are similar.

For simplicity, use z' for z in $H(u, \vec{x})$ to have $H(u, \vec{x}) = \forall z' \leq s(\vec{x})\ G(u, \vec{x}, z')$ and $H(u+1, \vec{x}) = \forall z \leq s(\vec{x})\ G(u+1, \vec{x}, z)$. Using the PV-reduction program P_u, we write an algorithm in $\square_{k-1}^p$ to find z' from z. W.l.o.g we can assume that the program begins with reading z. The algorithm M_u is defined as the following: Begin with the sequent

$$S(P_u^{<1}) = \forall z' \leq s(\vec{x}) G(u, \vec{x}, z') \Rightarrow G(u+1, \vec{x}, z)$$

Check the truth value of $z \leq s(\vec{x}) \rightarrow G(u+1, \vec{x}, z)$. If it is true, halt and answer 0. If not, follow the program in a way that all added simpler formulas to the left hand-side (right hand-side) of $S(P_u^{<1})$ becomes true (false). More precisely, at the stage m of the program, if P_m is the instruction [**Read** $X \leq t(\vec{x})$] and if the formula $\forall X \leq t(\vec{x}) C(X)$ that is occurred in the right hand-side of $S(P^{<m})$ is false ($\exists X \leq t(\vec{x}) C$ in the left hand-side is true), find X such that $X \leq t(\vec{x})$ and $C(X)$ becomes false (true). If not, continue. For the instruction [**Compute** Y **by** $r \leq t(\vec{x})$], if $Y = z'^k$ for some k and $t(\vec{x}) = s(\vec{x})$, check if $G(u, \vec{x}, r)$ is true or false. If it is false then halt and answer r. If not, continue. If $Y \notin \{z'^k\}_{k \geq 0}$, then continue.

The algorithm definitely halts and finds $r(u, \vec{x}, z)$ such that both

$$z \leq s(\vec{x}) \rightarrow r(u, \vec{x}, z) \leq s(\vec{x})$$

and

$$z \leq s(\vec{x}) \rightarrow [G(u, \vec{x}, r(u, \vec{x}, z)) \rightarrow G(u+1, \vec{x}, z)]$$

become valid. The reason is simple. If the algorithm does not find such an r, it must reach the end of the program. Based on our construction, all added

formulas to the right hand-side is false and all added formulas in the left hand-side is true. But there exists a quantifier-free subsequent of $S(P_u^{\leq l+1})$ such that $\mathrm{PV} \vdash S'$. Since S' consists of quantifier-free formulas, it should consist of added simpler formulas which implies that the left hand-side of S' is true while its right hand-side is false. Hence, S' is false. Therefore, the algorithm halts. But if it halts, there are two possibilities, either at the first stage $G(u+1, \vec{x}, z)$ is false, or in some stage there exists an r such that $G(u, \vec{x}, r(u, \vec{x}, z))$ is false. In both cases we have

$$z \leq s(\vec{x}) \rightarrow [G(u, \vec{x}, r(u, \vec{x}, z)) \rightarrow G(u+1, \vec{x}, z)].$$

We also have

$$z \leq s(\vec{x}) \rightarrow r(u, \vec{x}, z) \leq s(\vec{x})$$

In the first case because the output of the algorithm is 0 and in the second case, because we faced the instruction [**Compute** Y **by** $r \leq s(\vec{x})$] whose definition implies the claimed bound. Hence, the claim follows.

Now we show that the algorithm M_u computes a function in $\Box_k^p$. Note that the algorithm begins with checking $z \leq s(\vec{x}) \rightarrow G(u+1, \vec{x}, z)$ which is in Σ_{k-1}^b. Then in each stage of the reduction program, if the instruction is [**Read** $X \leq t(\vec{x})$], the algorithm checks the truth value of an existential sub-formula of $\forall z' \leq s(\vec{x}) G(u, \vec{x}, z')$ or a universal sub-formula of $G(u+1, \vec{x}, z)$ which implies that the formula is in Σ_{k-1}^b. And finally if the instruction is [**Compute** Y **by** $r \leq t(x)$], at the worst case, we have to check $G(u, \vec{x}, r)$ which is also in Σ_{k-1}^b. Hence, the algorithm is a constant number of Σ_{k-1}^b oracle questions and thus is in $\Box_k^p$.

Now let us investigate how complex this halting argument is. Since the length of the reduction program is a constant and

$$\mathrm{PV} \vdash \forall u < p(|\vec{x}|)\ H(u, \vec{x}) \rightarrow H(u+1, \vec{x})$$

it is easy to formalize the above mentioned argument in PV to show the following formula $(*)$:

$$\forall u < p(|\vec{x}|) \forall z \leq s(\vec{x}) [\exists w Com(\vec{x}, z, M_u, w) \rightarrow [G(u, \vec{x}, out(w)) \rightarrow G(u+1, \vec{x}, z)]]$$

and

$$\forall u < p(|\vec{x}|) \forall z \leq s(\vec{x}) [\exists w Com(\vec{x}, z, M_u, w) \rightarrow out(w) \leq s(\vec{x})] \quad (**)$$

Now apply the same argument for the PV-reduction programs N and K to have:

(i) $\forall z \leq s(\vec{x}) [\exists w Com(\vec{x}, z, N, w)) \rightarrow [(out(w) \leq t(\vec{x}) \rightarrow \neg A(\vec{x}, out(w))) \rightarrow G(0, \vec{x}, z)]]$.

(ii) $\exists w Com(\vec{x}, K, w) \rightarrow [G(p(|\vec{x}|), \vec{x}, out(w)) \rightarrow \bot]$.

(iii) $\exists w Com(\vec{x}, K, w) \rightarrow out(w) \leq s(\vec{x})$

Now define the algorithm M as running K on 0, then put M_u's end to end beginning from $u = p(|\vec{x}|)$ till $u = 0$ and at last run N. We claim that this M works. First note that M is a result of polynomially many computational steps in $\square_k^p$ and hence it is also in $\square_k^p$. Secondly note that by the length induction in PV on $p(|\vec{x}|) \dot{-} u$ and using $(*)$ and $(**)$ we can prove

$$\mathrm{PV} \vdash \exists w Com(\vec{x}, M, w) \to \forall u < p(|\vec{x}|)\ out(w_u) \leq s(\vec{x}).$$

and

$$\mathrm{PV} \vdash \exists w Com(\vec{x}, M, w) \to \forall u < p(|\vec{x}|) \neg G(u, \vec{x}, out(w_u)).$$

Hence

$$\mathrm{PV} \vdash \exists w Com(\vec{x}, M, w) \to [out(w) \leq t(\vec{x}) \wedge A(\vec{x}, out(w))]$$

which complete the proof. □

2.4 Deterministic Flows

In this section we will develop a theory for deterministic reductions and deterministic flows. We will use this theory to transform any proof in a bounded theory of arithmetic to a term-length sequence of provably simple game reductions. This technique provides a combinatorial characterization for the bounded consequences of bounded theories of arithmetic, including the interesting case of higher search problems of the theories IU_k and T_n^k and low complexity consequences of stronger theories $I\Delta_0(\exp)$, PRA and $\mathrm{PA} + \mathrm{TI}(\alpha)$. Our characterization also presents another proof for the previous characterizations of low complexity consequences of the theory T_2^k appeared in [19], [21] and [6].

2.4.1 Reductions and Flows

In the following, we will define the notion of a deterministic reduction as the building block of the followed deterministic flows. These reductions are the generalization of the usual polynomial-time reductions between total NP search problems and the deterministic reductions between k-turn games as introduced in the Section 2.2.

Definition 2.4.1. *Let A be a formula whose negative sub-formulas are all atomic. By the σ-prenex form of A, we means the result of the following process: First bring out all existential quantifiers, then all universal quantifiers and so on. If we begin by the universal quantifiers, it is called π-prenex form of A.*

Definition 2.4.2. *Let $\alpha \in \{\sigma, \pi\}$ and $A(\vec{x})$ and $B(\vec{x})$ be some bounded formulas in the α-prenex form with at most k alternations of quantifiers, $\{\vec{F}_i\}_{i=1}^k$ be a sequence of sequences of terms and $\mathcal{B} \supseteq \mathcal{R}$ a theory. By recursion on k, we will define $F = \{\vec{F}_i\}_{i=1}^k$ as a deterministic α-reduction, from $B(\vec{x})$ to $A(\vec{x})$ and we will denote it by $A(\vec{x}) \geq_\alpha^{\mathcal{B},F} B(\vec{x})$ when:*

(*i*) *If A, B are quantifier-free, a sequence of sequences of terms is both a σ- and a π-deterministic reduction from B to A iff $\mathcal{B} \vdash A(\vec{x}) \to B(\vec{x})$.*

(*ii*) *If* $\alpha = \pi$, *we have* $A = \forall \vec{u} C(\vec{x}, \vec{u})$, $B = \forall \vec{v} D(\vec{x}, \vec{v})$ *where the universal quantifiers are the whole block of left-most universal quantifiers (possibly empty) and* $F = \{\vec{F}_i\}_{i=1}^{k+1}$ *is a sequence of terms, then* $A(\vec{x}) \geq_\pi^{\mathcal{B},F} B(\vec{x})$ *iff*

$$C(\vec{x}, \vec{F}_{k+1}(\vec{x}, \vec{v})) \geq_\sigma^{\mathcal{B},\hat{F}} D(\vec{x}, \vec{v})$$

where $\hat{F} = \{\vec{F}_i\}_{i=1}^{k}$.

(*iii*) *If* $\alpha = \sigma$, *we have* $A = \exists \vec{u} C(\vec{x}, \vec{u})$, $B = \exists \vec{v} D(\vec{x}, \vec{v})$ *where the existential quantifiers are the whole block of left-most existential quantifiers (possibly empty) and* $F = \{\vec{F}_i\}_{i=1}^{k+1}$ *is a sequence of terms, then* $A(\vec{x}) \geq_\sigma^{\mathcal{B},F} B(\vec{x})$ *iff*

$$C(\vec{x}, \vec{u}) \geq_\pi^{\mathcal{B},\hat{F}} D(\vec{x}, \vec{F}_{k+1}(\vec{x}, \vec{u}))$$

where $\hat{F} = \{\vec{F}_i\}_{i=1}^{k}$.

It is possible to extend the definition to all bounded formula $A(\vec{x})$ *and* $B(\vec{x})$ *whose negative sub-formulas are all atomic, in the following way: We say* $F = \{\vec{F}_i\}_{i=1}^{k}$ *is a deterministic* α*-reduction, from* $B(\vec{x})$ *to* $A(\vec{x})$ *iff* $F = \{\vec{F}_i\}_{i=1}^{k}$ *is a deterministic* α*-reduction, from* $\tilde{B}(\vec{x})$ *to* $\tilde{A}(\vec{x})$, *where* $\tilde{A}(\vec{x})$ *and* $\tilde{B}(\vec{x})$ *are the* α*-prenex forms of* A *and* B, *respectively.*

Finally, we say B *is* $(\pi, \mathcal{B})$*-deterministicly reducible to* A *and we write* $A \geq_\pi^{\mathcal{B}} B$, *when there exists a sequence of sequences of terms* F *such that* $A \geq_\pi^{\mathcal{B},F} B$. *Moreover, by the equivalence* $A \equiv_\pi^{\mathcal{B},E,F} B$ *we mean the conjunction of* $A \geq_\pi^{\mathcal{B},E} B$ *and* $B \geq_\pi^{\mathcal{B},F} A$ *and we define* $(\sigma, \mathcal{B})$*-deterministic reducibility and equivalence dually by replacing* π *to* σ *everywhere. Note that whenever the theory* $\mathcal{B}$ *is clear from the context, we drop it from the superscripts everywhere.*

Example 2.4.3. *In this example we will draw the reader's attention to the difference between* π*- and* σ*-reductions. Consider the formula*

$$A = \forall y \leq t(x) B(x, y) \vee \exists z \leq t(x) \neg B(x, z)$$

where B *is quantifier-free. Working with* π*-reductions, it is clear that we have* $\top \geq_\pi A$ *because we can first read* y *and then witness* z *by* y. *But if we work with the* σ*-reductions, the order of the variables changes and we have to witness* z *first without the knowledge of the value* y *which is clearly impossible in the general setting.*

Convention. From now on we will assume that wherever we talk about the deterministic reductions, the language and the base theory has the following properties:

(*i*) There exists a subset of monotone majorizing terms provably in $\mathcal{B}$, i.e. there exists a set of terms M such that for any term $t(\vec{x})$ there exists $s(\vec{x}) \in M$ such that $\mathcal{B} \vdash t(\vec{x}) \leq s(\vec{x})$ and for any $r(\vec{x}) \in M$, $\mathcal{B} \vdash \vec{x} \leq \vec{y} \to r(\vec{x}) \leq r(\vec{x})$.

(*ii*) For any quantifier-free formula $A(\vec{x})$, there exists a term $t(\vec{x})$ such that $\mathcal{B} \vdash [t(\vec{x}) = 0 \to A(\vec{x})] \wedge [t(\vec{x}) \neq 0 \to \neg A(\vec{x})]$. We call this term the characteristic function for the formula $A(\vec{x})$.

Definition 2.4.4. *Let $A(\vec{x}), B(\vec{x}) \in \Pi_k$ be two formulas and $\alpha \in \{\sigma, \pi\}$. A $(\Pi_k, \mathcal{B}, \alpha)$-deterministic flow from $A(\vec{x})$ to $B(\vec{x})$ is the following data: A term $t(\vec{x})$, a formula $H(u, \vec{x}) \in \Pi_k$ and sequences of terms E_0, E_1, G_0, G_1 and $F(u)$ such that the following statements are provable in $\mathcal{B}$:*

(i) $H(0, \vec{x}) \equiv_\alpha^{(E_0, E_1)} A(\vec{x})$.

(ii) $H(t(x), \vec{x}) \equiv_\alpha^{(G_0, G_1)} B(\vec{x})$.

(iii) $\forall u < t(x) H(u, \vec{x}) \geq_\alpha^{F(u)} H(u+1, \vec{x})$.

If there exists a deterministic $(\Pi_k, \mathcal{B}, \alpha)$-flow from $A(\vec{x})$ to $B(\vec{x})$ we will write $A(\vec{x}) \rhd_{d,\alpha}^{(\Pi_k, \mathcal{B})} B(\vec{x})$. Moreover, if Γ and Δ are sequents of formulas in Π_k, by $\Gamma \rhd_{d,\alpha}^{(\Pi_k, \mathcal{B})} \Delta$ we mean $\bigwedge \Gamma \rhd_{d,\alpha}^{(\Pi_k, \mathcal{B})} \bigvee \Delta$. The case for $(\Sigma_k, \mathcal{B}, \alpha)$-flows is defined similarly by replacing Π_k with Σ_k.

For the sake of brevity, let us have some convention on notation:

Convention. In the remaining parts of this section, whenever we use the letters Φ and α, we mean $\Phi \in \{\Sigma_k, \Pi_k\}$ for some k and $\alpha \in \{\sigma, \pi\}$. In cases that k has been already fixed and the type of the class can be chosen from Σ and Π, we write Φ_k. In some cases, it is important to have the same type of α as Φ, i.e., either $[\Phi = \Sigma_k$ for some k and $\alpha = \sigma]$ or $[\Phi = \Pi_k$ for some k and $\alpha = \pi]$. In these cases, we use the letter ϕ instead of α to mention the dependency between Φ and ϕ. Moreover, when we work with a fixed choice for Φ, α and $\mathcal{B}$, we use the shorthand $\rhd$ for $\rhd_{d,\alpha}^{(\Phi, \mathcal{B})}$. However, sometimes, we put some parts of the pairs $(\Phi, \mathcal{B})$ and (d, α) for $\rhd$ to emphasize on those specific parts. For instance, if we write $\rhd_\alpha^\Phi$, we want to emphasize on the class of the flow and the fact that it is a α-flow.

2.4.2 The Main Theorem

Now were are ready to state the main theorem of this section. The theorem relates the provability of bounded formulas in bounded theories of arithmetic to the existence of a uniform term-length sequence of deterministic reductions. The latter can also be interpreted as the existence of a uniform term-length sequence of games with a uniform term-based sequence of methods to transfer the winning strategies along them.

Theorem 2.4.5. *(Main Theorem) Assume $\Gamma(\vec{x}) \cup \Delta(\vec{x}) \subseteq \Phi_k$ and $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathfrak{B}(\Phi_k, \mathcal{A})$ then $\mathfrak{B}(\Phi_k, \mathcal{A}) \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$ iff $\Gamma \rhd_\phi^{(\Phi_k, \mathcal{B})} \Delta$.*

In the following we will prove a sequence of lemmas to make a high-level calculus for deterministic flows. Then we will use this calculus to show that this flow interpretation is sound and complete with respect to the corresponding bounded arithmetic as stated in Theorem 2.4.5. All lemmas are true both for both π and σ deterministic flows.

Lemma 2.4.6. *(Conjunction Application) Let $C(\vec{x}) \in \Phi$ be a formula. If $A(\vec{x}) \rhd^\Phi B(\vec{x})$ then $A(\vec{x}) \wedge C(\vec{x}) \rhd^\Phi B(\vec{x}) \wedge C(\vec{x})$.*

Proof. Since $A(\vec{x}) \rhd B(\vec{x})$, by Definition 2.4.4, there exists a term $t(\vec{x})$, a formula $H(u,\vec{x}) \in \Phi$ and sequences of terms E_0, E_1, G_0, G_1 and $F(u)$ such that

$$\mathcal{B} \vdash A(\vec{x}) \equiv^{E_0,E_1} H(0,\vec{x}),$$

$$\mathcal{B} \vdash B(\vec{x}) \equiv^{G_0,G_1} H(t(\vec{x}),\vec{x}),$$

and

$$\mathcal{B} \vdash \forall u < t(\vec{x})\ H(u,\vec{x}) \geq^{F(u)} H(u+1,\vec{x}).$$

Now define $t' = t$, $H'(u,\vec{x}) = H(u,\vec{x}) \wedge C(\vec{x})$ and E_0', E_1', G_0', G_1' and $F'(u)$ as the corresponding sequences of terms extending their counterparts by using the quantifiers in C to witness themselves by the identity terms. It is clear that the new data is a deterministic $(\Phi,\mathcal{B},\alpha)$-flow from $A(\vec{x}) \wedge C(\vec{x})$ to $B(\vec{x}) \wedge C(\vec{x})$. □

Lemma 2.4.7. *(Disjunction Application) Let $C(\vec{x}) \in \Phi$ be a formula. If $A(\vec{x}) \rhd^{\Phi} B(\vec{x})$ then $A(\vec{x}) \vee C(\vec{x}) \rhd^{\Phi} B(\vec{x}) \vee C(\vec{x})$.*

Proof. Since $A(\vec{x}) \rhd B(\vec{x})$ then by Definition 2.4.4, there exists a term $t(\vec{x})$, a formula $H(u,\vec{x}) \in \Phi$ and sequences of terms E_0, E_1, G_0, G_1 and $F(u)$ such that the conditions in the Definition 2.4.4 is provable in $\mathcal{B}$. Now define $t' = t$, $H'(u,\vec{x}) = H(u,\vec{x}) \vee C(\vec{x})$ and E_0', E_1', G_0', G_1' and $F'(u)$ as the corresponding sequences of terms extending their counterparts by using the quantifiers in C to witness themselves by the identity terms. It is clear that the new data is a deterministic $(\Phi,\mathcal{B},\alpha)$-flow from $A(\vec{x}) \vee C(\vec{x})$ to $B(\vec{x}) \vee C(\vec{x})$. □

Lemma 2.4.8. *(i) (Weak Gluing) If $A(\vec{x}) \rhd B(\vec{x})$ and $B(\vec{x}) \rhd C(\vec{x})$ then $A(\vec{x}) \rhd C(\vec{x})$.*

(ii) (Strong Gluing) If $A(y,\vec{x}) \rhd A(y+1,\vec{x})$ then $A(0,\vec{x}) \rhd A(s,\vec{x})$.

Proof. For (i), since $A(\vec{x}) \rhd B(\vec{x})$ there exists a term $t(\vec{x})$, a formula $H(u,\vec{x}) \in \Phi$ and sequences of terms E_0, E_1, G_0, G_1 and $F(u)$ such that $\mathcal{B}$ proves the conditions in the Definition 2.4.4. On the other hand since $B(\vec{x}) \rhd C(\vec{x})$ we have the corresponding data for $B(\vec{x})$ to $C(\vec{x})$ which we denote $t'(\vec{x})$, $H'(u,\vec{x})$, E_0', E_1', G_0', G_1' and $F'(u)$. Define $r(\vec{x}) = t(\vec{x}) + t'(\vec{x}) + 1$,

$$I(u,\vec{x}) = \begin{cases} H(u,\vec{x}) & u \leq t(\vec{x}) \\ B(\vec{x}) & u = t(\vec{x}) + 1 \\ H'(u \dot{-} t(\vec{x}) \dot{-} 2,\vec{x}) & t(\vec{x}) + 1 < u \leq t(\vec{x}) + t'(\vec{x}) + 1 \end{cases}$$

and the sequence of terms in the same pointwise way. Then, it is easy to check that this new data is a deterministic $(\Phi,\mathcal{B},\alpha)$-flow from $A(\vec{x})$ to $C(\vec{x})$.

For (ii), if we have $A(y,\vec{x}) \rhd A(y+1,\vec{x})$ it is enough to glue all copies of the sequences of reductions for $0 \leq y \leq s$, to have $A(0,\vec{x}) \rhd A(s,\vec{x})$. More precisely, assume that all reductions have the same length $t'(\vec{x})$ greater than $t(s,\vec{x})$. This is an immediate consequence of the facts that we can find a monotone majorization for $t(y,\vec{x})$ like $r(y,\vec{x})$, and since $y \leq s$ we have $t(y,\vec{x}) \leq r(y,\vec{x}) \leq r(s,\vec{x})$. Now it is enough to repeat the last formula in the flow to make the flow longer to reach

the length $t'(\vec{x},\vec{z}) = r(s,\vec{x})$ where $\vec{z}$ is a vector of variables in s. Now, define $t''(\vec{x},\vec{z}) = s \times (t'(\vec{x})+2)$,

$$I(u,\vec{x}) = \begin{cases} H(u,y,\vec{x}) & y(t'+2) < u < (y+1)(t'+2) \\ A(y,\vec{x}) & u = y(t'+2) \end{cases}$$

and

$$F(u) = \begin{cases} F(u,y) & y(t'+2) < u < (y+1)(t'+2) \dot{-} 1 \\ E_0(u,y) & u = y(t'+2) \\ G_1(u,y+1) & u = (y+1)(t'+2) \dot{-} 1 \end{cases}$$

and $E'_0 = E'_1 = G'_0 = G'_1 = id$. It is easy to see that this new sequence is a deterministic $(\Phi,\mathcal{B},\alpha)$-flow from $A(0,\vec{x})$ to $A(s,\vec{x})$. □

Lemma 2.4.9. *(Quantifier Application)*

(*i*) *If* $A(\vec{x},y) \rhd_\pi^{\Pi_k} B(\vec{x},y)$ *then for any* $\alpha \in \{\sigma,\pi\}$, $\forall y \le t(\vec{x})A(\vec{x},y) \rhd_\alpha^{\Pi_k} \forall y \le t(\vec{x})B(\vec{x},y)$.

(*ii*) *If* $A(\vec{x},y) \rhd_\sigma^{\Sigma_k} B(\vec{x},y)$ *then for any* $\alpha \in \{\sigma,\pi\}$, $\exists y \le t(\vec{x})A(\vec{x},y) \rhd_\alpha^{\Sigma_k} \exists y \le t(\vec{x})B(\vec{x},y)$.

Proof. For (*i*), since $A(\vec{x},y) \rhd_\pi^{\Pi_k} B(\vec{x},y)$, there exists a term $s(\vec{x},y)$, a formula $H(u,\vec{x},y) \in \Pi_k$ and sequences of sequences of terms E_0, E_1, G_0, G_1 and $F(u)$ such that the conditions of the Definition 2.4.2 are provable in $\mathcal{B}$. W.l.o.g we can assume that s is monotone, because any term is majorizable by a monotone term and we can extend the sequence by repeating the last formula to reach that majorization as the length of the flow. Define $t'(\vec{x})$ as $s(\vec{x},t(\vec{x}))$ and $H'(u,\vec{x}) = \forall y \le t(\vec{x})I(u,\vec{x},y)$ where

$$I(u,\vec{x},y) = \begin{cases} H(u,y,\vec{x}) & u \le s(\vec{x},y) \\ H(s(\vec{x},y),y,\vec{x}) & \text{o.w.} \end{cases}$$

and

$$C_u = \begin{cases} F_u & u+1 \le s(\vec{x},y) \\ Id & \text{o.w.} \end{cases}$$

Firstly, it is clear that $H'(0,\vec{x}) \equiv_\pi \forall y \le tA(\vec{x},y)$ by the reductions which read y and witness it by itself and then apply the reductions E_0 and E_1. The reason is that $0 \le s(\vec{x},y)$ and hence $I(0,\vec{x},y) = H(0,y,\vec{x})$.

Secondly, note that $H'(u,\vec{x}) \ge_\pi H'(u+1,\vec{x})$ by witnessing the outmost quantifier $\forall y$ by itself and then applying C_u. Thirdly, $H'(t'(\vec{x}),\vec{x}) \equiv_\pi \forall y \le tB(\vec{x},y)$ by the reductions which read y and witness it by itself and then apply the reductions G_0 and G_1. To prove this claim, first note that we can assume $y \le t$, because otherwise, both sides of the reduction will be false regardless of the reduction. Then using $y \le t$ and the monotonicity of s we have $s(\vec{x},y) \le t'(\vec{x})$ which implies $I(t'(\vec{x}),\vec{x},y) = I(s(\vec{x},y),y,\vec{x})$ and since $I(s(\vec{x},y),y,\vec{x}) = H(s(\vec{x},y),y,\vec{x})$ is π-equivalent to B by the reductions G_0, G_1, the claim follows. Finally note that all the formulas in the flow begin with a universal quantifier, therefore, we can also claim that all the reductions are σ-reductions and hence the flow is also a σ-flow. The proof of (*ii*) is similar. □

Lemma 2.4.10. *(Conjunction and Disjunction Rules)*

(*i*) *If* $\Gamma, A \rhd \Delta$ *or* $\Gamma, B \rhd \Delta$ *then* $\Gamma, A \wedge B \rhd \Delta$.

(*ii*) *If* $\Gamma_0 \rhd \Delta_0, A$ *and* $\Gamma_1 \rhd \Delta_1, B$ *then* $\Gamma_0, \Gamma_1 \rhd \Delta_0, \Delta_1, A \wedge B$.

(*iii*) *If* $\Gamma \rhd \Delta, A$ *or* $\Gamma \rhd \Delta, B$ *then* $\Gamma \rhd \Delta, A \vee B$.

(*iv*) *If* $\Gamma_0, A \rhd \Delta_0$ *and* $\Gamma_1, B \rhd \Delta_1$ *then* $\Gamma_0, \Gamma_1, A \vee B \rhd \Delta_0, \Delta_1$.

Proof. (*i*) and (*iii*) are trivial simply because firstly we have $A \wedge B \geq A$, $A \wedge B \geq B$, $A \geq A \vee B$ and $B \geq A \vee B$ by using the quantifiers in both sides to witness themselves and forget the irrelevant parts and then add the needed formula in the beginning or the end of the flow. For (*ii*), if $\Gamma_0 \rhd \Delta_0, A$, then by conjunction application with $\bigwedge \Gamma_1$ we have $\bigwedge \Gamma_0 \wedge \bigwedge \Gamma_1 \rhd (\bigvee \Delta_0 \vee A) \wedge \bigwedge \Gamma_1$. Moreover, we have $\bigwedge \Gamma_1 \rhd \bigvee \Delta_1 \vee B$ and again by conjunction application $\bigwedge \Gamma_1 \wedge (\bigvee \Delta_0 \vee A) \rhd (\bigvee \Delta_1 \vee B) \wedge (\bigvee \Delta_0 \vee A)$. Therefore by weak gluing

$$\bigwedge \Gamma_0 \wedge \bigwedge \Gamma_1 \rhd (\bigvee \Delta_1 \vee B) \wedge (\bigvee \Delta_0 \vee A).$$

But it is easy to see that

$$(\bigvee \Delta_1 \vee B) \wedge (\bigvee \Delta_0 \vee A) \geq \bigvee \Delta_1 \vee \bigvee \Delta_0 \vee (A \wedge B).$$

Hence

$$\Gamma_0, \Gamma_1 \rhd \Delta_0, \Delta_1, (A \wedge B).$$

For (*iv*), if $\Gamma_0, A \rhd \Delta_0$ then by disjunction application with $\bigwedge \Gamma_1 \wedge B$ we have

$$(\bigwedge \Gamma_0 \wedge A) \vee (\bigwedge \Gamma_1 \wedge B) \rhd \bigvee \Delta_0 \vee (\bigwedge \Gamma_1 \wedge B).$$

Moreover, we have $\bigwedge \Gamma_1 \wedge B \rhd \bigvee \Delta_1$, hence again by disjunction application

$$(\bigwedge \Gamma_1 \wedge B) \vee \bigvee \Delta_0 \rhd \bigvee \Delta_0 \vee \bigvee \Delta_1.$$

Hence, by weak gluing,

$$(\bigwedge \Gamma_0 \wedge A) \vee (\bigwedge \Gamma_1 \wedge B) \rhd \bigvee \Delta_0 \vee \bigvee \Delta_1.$$

However, it is clear that

$$\bigwedge \Gamma_0 \wedge \bigwedge \Gamma_1 \wedge (A \vee B) \geq (\bigwedge \Gamma_0 \wedge A) \vee (\bigwedge \Gamma_1 \wedge B).$$

Hence,

$$\Gamma_0, \Gamma_1, (A \vee B) \rhd \Delta_0, \Delta_1.$$

□

In the following, wherever we write $\neg A$, we mean the statement B resulting from pushing the negation inside to the level of atomic formulas.

The following lemma provides a machinery to compute the value of the formula $A \in \Phi_k \in \{\Pi_k, \Sigma_k\}$ by a deterministic $(\Sigma_{k+1}, \mathcal{B}, \alpha)$-flow of reductions for any $\alpha \in \{\pi, \sigma\}$. This is a very important tool to reduce the complexity of deciding a complex formula to just deciding one equality. We will see its use in full force in the case of handling the contraction rule.

Lemma 2.4.11. *(Computability of the characteristic functions) Suppose $\mathcal{B}$ has a characteristic term for any quantifier-free formula then for any $\alpha \in \{\pi, \sigma\}$ and any $\Phi \in \{\Pi_k, \Sigma_k\}$ if $A(\vec{x}) \in \Phi$ then we have:*

$$\triangleright_\alpha^{(\Sigma_{k+1},\mathcal{B})} \exists i \leq 1\ [(i = 1 \to A) \wedge (i = 0 \to \neg A)]$$

Proof. We say a bounded quantifier is constant if it has the form $\forall z \leq s(z = s \to D(z))$ or $\exists z \leq t(z = s \wedge D(z))$ for some term s. We denote these quantifiers by $\forall\{z = s\}$ and $\exists\{z = s\}$. To prove the theorem, use induction on the sum of the number of non-constant quantifiers of A and the number of disjunctions and conjunctions of A.

If all the quantifiers in A are constant, then it is enough to first eliminate all the quantifiers in A by substituting the variables by the constant terms that the constant quantifiers suggest, i.e., substituting the variable z in the quantifier $Q\{z = s\}$ by s. Call this quantifier-free formula B and put $i = \chi_B$. If we witness all the essentially existential quantifiers by the terms that they suggest, then we reach the implication

$$(\chi_B = 1 \to B) \wedge (\chi_B = 0 \to \neg B)$$

which is provable in $\mathcal{B}$ by the assumption.

If $A = \vec{Q}\{\vec{z} = \vec{s}\}(B \wedge C)$ where $Q_n \in \{\forall, \exists\}$, then by IH,

$$\triangleright_\alpha^{(\Sigma_{k+1},\mathcal{B})} \exists j \leq 1\ [(j = 1 \to \vec{Q}\{\vec{z} = \vec{s}\}B) \wedge (j = 0 \to \neg\vec{Q}\{\vec{z} = \vec{s}\}B)]$$

and

$$\triangleright_\alpha^{(\Sigma_{k+1},\mathcal{B})} \exists k \leq 1\ [(k = 1 \to \vec{Q}\{\vec{z} = \vec{s}\}C) \wedge (k = 0 \to \neg\vec{Q}\{\vec{z} = \vec{s}\}C)].$$

On the other hand, it is possible to reduce

$$\exists i \leq 1\ [(i = 1 \to \vec{Q}\{\vec{z} = \vec{s}\}(B \wedge C)) \wedge (i = 0 \to \neg\vec{Q}\{\vec{z} = \vec{s}\}(B \wedge C))]$$

to the conjunction of two statements

$$\exists j \leq 1\ [(j = 1 \to \vec{Q}\{\vec{z} = \vec{s}\}B) \wedge (j = 0 \to \neg\vec{Q}\{\vec{z} = \vec{s}\}B)]$$

and

$$\exists k \leq 1\ [(k = 1 \to \vec{Q}\{\vec{z} = \vec{s}\}C) \wedge (k = 0 \to \neg\vec{Q}\{\vec{z} = \vec{s}\}C)].$$

To prove this, witness i by jk, the quantifiers in $\vec{Q}$ by the terms that they suggest and the other quantifiers with themselves. Therefore, by conjunction application and then gluing, we have

$$\triangleright_\alpha^{(\Sigma_{k+1},\mathcal{B})} \exists i \leq 1\ [(i = 1 \to \vec{Q}\{\vec{z} = \vec{s}\}(B \wedge C)) \wedge (i = 0 \to \neg\vec{Q}\{\vec{z} = \vec{s}\}(B \wedge C))].$$

The case for disjunction is similar to the conjunction case.

If $A = \vec{Q}\{\vec{v} = \vec{s}\}\forall z \leq t(\vec{x})B(\vec{x}, z)$ where $Q_n \in \{\forall, \exists\}$, then define $G(u)$ as

$$\exists k \leq 1\ [(k = 1 \to \vec{Q}\tilde{B}(\vec{x}, u)) \wedge (k = 0 \to \neg\vec{Q}\tilde{B}(\vec{x}, u))].$$

where $\tilde{B}(\vec{x}, u)$ is $\forall\{z = u\}B(\vec{x}, z)$ and $\vec{Q}$ stands for $\vec{Q}\{\vec{v} = \vec{s}\}$. By IH we have a $(\Sigma_{k+1}, \mathcal{B}, \alpha)$-flow from $\top$ to $G(u+1)$ which is

$$\exists k \leq 1\ [(k = 1 \to \vec{Q}\tilde{B}(\vec{x}, u+1)) \wedge (k = 0 \to \neg\vec{Q}\tilde{B}(\vec{x}, u+1))]$$

Define $H(u)$ as

$$\exists i \leq 1\ [(i = 1 \to \vec{Q}\forall z \leq u\ B(\vec{x}, z)) \wedge (i = 0 \to \neg\vec{Q}\forall z \leq u\ B(\vec{x}, z))].$$

Now, we want to prove the existence of a reduction from $H(u+1)$ which is

$$\exists j \leq 1\ [(j = 1 \to \vec{Q}\forall z \leq u+1\ B(\vec{x}, z)) \wedge (j = 0 \to \neg\vec{Q}\forall z \leq u+1\ B(\vec{x}, z))].$$

to the conjunction of $G(u+1)$ and $H(u)$. For this purpose, witness j by ik. Then for the other quantifiers use the following scheme: Note that we have three possible cases, the case when $i = k = 1$, the case $i = 1$, $k = 0$ and the case $i = 0$. In each case, some parts of the formulas, will be true regardless of the reduction that we will present. Hence, we ignore them altogether and we call the other formulas the main formulas.

Now, if $i = k = 1$, then the main formulas are $\vec{Q}\forall z \leq u\ B(\vec{x}, z)$, $\tilde{B}(\vec{x}, u+1)$ and $\vec{Q}\forall z \leq u+1\ B(\vec{x}, z)$. To reduce $\vec{Q}\forall z \leq u+1\ B(\vec{x}, z)$ to the conjunction of $\vec{Q}\forall z \leq u\ B(\vec{x}, z)$ and $\vec{Q}\tilde{B}(\vec{x}, u+1)$, first witness constant quantifiers in $\vec{Q}$ by the terms that they suggest. Then read $z \leq u+1$, if $z = u+1$ use $\tilde{B}(\vec{x}, u+1)$ and $\forall z \leq u+1\ B(\vec{x}, z)$ as the main formulas and ignore $\forall z \leq u\ B(\vec{x}, z)$. Then witness the last universal quantifier of $\tilde{B}(\vec{x}, u+1)$ by $u+1$ and all the other quantifiers in $\forall z \leq u+1\ B(\vec{x}, z)$ and $\tilde{B}(\vec{x}, u+1)$ with themselves. If $z < u+1$, then use $\forall z \leq u\ B(\vec{x}, z)$ and $\forall z \leq u+1\ B(\vec{x}, z)$ as the main formulas and again witness everything with themselves. If $i = 1$ and $k = 0$, then use $\neg\vec{Q}\forall z \leq u+1\ B(\vec{x}, z)$ and $\neg\vec{Q}\tilde{B}(\vec{x}, u+1)$ as the main formulas and witness constant quantifiers in $\vec{Q}$ by the terms that they suggest. Then use $u+1$ for z and witness all the variables with themselves. Finally if $i = 0$, then use $\neg\vec{Q}\forall z \leq u\ B(\vec{x}, z)$ and $\neg\vec{Q}\forall z \leq u+1\ B(\vec{x}, z)$ as the main formulas and witness constant quantifiers in $\vec{Q}$ by the terms that they suggest and all the other variables with themselves.

Therefore $G(u+1) \wedge H(u) \rhd H(u+1)$. By IH, $\rhd G(u+1)$. Hence, by conjunction application $H(u) \rhd G(u+1) \wedge H(u)$ and then by gluing $H(u) \rhd H(u+1)$ and finally by strong gluing $H(0) \rhd H(t(\vec{x}))$. Since $H(0) \equiv G(0)$ and $\rhd G(0)$, hence $\rhd H(0)$ which means $\rhd H(t(\vec{x}))$.

The case $A = \vec{Q}\{\vec{v} = \vec{s}\}\exists z \leq t(\vec{x})B(\vec{x}, z)$ is similar to the universal case. □

Lemma 2.4.12. *(Canonical Normal Form) For any formula $C(\vec{x}) \in \Sigma_{k+1}$, there exists $\tilde{C}(\vec{x}, \vec{u}) \in \Pi_k$ and some terms $\vec{s}$ such that the σ-prenex form of $\tilde{C}(\vec{x}, \vec{u})$ is quantifier-free or it begins with a universal quantifier and $\exists \vec{u} \leq \vec{s}\ \tilde{C}(\vec{x}, \vec{u})$ is σ-deterministic equivalent to C. The same also holds for universal quantifiers, Π_{k+1} and π-equivalence.*

Proof. We will prove the claim by induction on the complexity of C. If C is quantifier-free, then pick $\tilde{C} = C$ and pick $\vec{u}$ as the empty vector. If C begins

with a universal formula then $\tilde{C} = C$ and pick $\vec{u}$ as the empty vector again. If $C = \exists y \leq t\ D$, then pick $\tilde{C} = \tilde{D}$ and add y to $\vec{u}$ and t to $\vec{s}$. For the cases $C = D \wedge E$ or $C = D \vee E$, since their proofs are similar, we will only check the disjunction case. If $C = D \vee E$, by IH, there exists $\tilde{C}$ and $\tilde{D}$ and $\vec{s}$ and $\vec{r}$ such that $C \equiv_\sigma \exists \vec{u} \leq \vec{s}\ \tilde{C}$ and $D \equiv_\sigma \exists \vec{v} \leq \vec{r}\ \tilde{D}$. Hence by propositional rules, it is clear that

$$C \vee D \equiv_\sigma \exists \vec{u} \leq \vec{s}\, \tilde{C} \vee \exists \vec{v} \leq \vec{r}\, \tilde{D}$$

But

$$\exists \vec{u_0} \leq \vec{s}\, \tilde{C} \vee \exists \vec{v_0} \leq \vec{r}\, \tilde{D} \equiv_\sigma \exists \vec{u_1} \leq \vec{s}\, \exists \vec{v_1} \leq \vec{r}\, (\tilde{C} \vee \tilde{D})$$

(For the moment we put some indices for the variables $\vec{u}$ and $\vec{v}$ for the referring purpose.) To show the latter, for both reductions, when we read an existential quantifier $w \in \vec{u} \cup \vec{v}$ with the bound p, if $w_i \leq p$ use w_i to witness w_{1-i}, if not just use zero. From right to left, if at least for one variable w_1 we have $w_1 > p$, then this choice for the variable w_1 makes the left hand-side of the reduction false, regardless the choice of the other variables, which implies the reduction. If for all the variables we have $w_1 \leq p$, then after using identity reduction both sides will be equal and there is nothing to prove. For the other direction, let $\vec{u_0}'$ and $\vec{v_0}'$ be the variables that do not meet their bounds in $\vec{u_0}$ and $\vec{v_0}$, respectively. If both $\vec{u_0}'$ and $\vec{v_0}'$ have some variables, as before, it makes both $\exists \vec{u_0} \leq \vec{s}\ \tilde{C}$ and $\exists \vec{v_0} \leq \vec{r}\ \tilde{D}$ false and hence we have the reduction. If $\vec{u_0}$ is non-empty and $\vec{v_0}$ is empty, then $\exists \vec{u_0} \leq \vec{s}\ \tilde{C}$ is false, regardless of the other parts of the reduction. Since we choose zero to witness the variables $\vec{u_1}'$, all $\vec{u_1}$ meet their bounds and therefore the reduction is reduced to the fact that the disjunction of $\tilde{D}$ and a substitution of $\tilde{C}$ is reducible to $\tilde{D}$. The proof for the other cases are similar. □

Lemma 2.4.13. *(Negation Rules) If* $\Gamma, \Delta \subseteq \Phi_{k+1}$ *and* $A \in \Pi_k \cup \Sigma_k$ *then*

(*i*) *If* $\Gamma, A \rhd_\phi^{\Phi_{k+1}} \Delta$ *then* $\Gamma \rhd_\phi^{\Phi_{k+1}} \Delta, \neg A$.

(*ii*) *If* $\Gamma \rhd_\phi^{\Phi_{k+1}} \Delta, A$ *then* $\Gamma, \neg A \rhd_\phi^{\Phi_{k+1}} \Delta$.

Proof. Since we have conjunction and disjunction application, it is enough to prove the claim:

Claim. If $A(\vec{x}) \in \Pi_k \cup \Sigma_k$, then

$(*)$ $\top \rhd_\phi^{\Phi_{k+1}} A(\vec{x}) \vee \neg A(\vec{x})$.

$(**)$ $A(\vec{x}) \wedge \neg A(\vec{x}) \rhd_\phi^{\Phi_{k+1}} \bot$.

The reason for this sufficiency is the following:

For (i), if we have $\Gamma, A \rhd \Delta$ then $\bigwedge \Gamma \wedge A \rhd \bigvee \Delta$, hence by disjunction application we have $(\bigwedge \Gamma \wedge A) \vee \neg A \rhd \bigvee \Delta \vee \neg A$. By the claim we have $\rhd A \vee \neg A$, therefore by conjunction application $\bigwedge \Gamma \rhd \bigwedge \Gamma \wedge (A \vee \neg A)$. But, it is easy to see that $\bigwedge \Gamma \wedge (A \vee \neg A) \geq (\bigwedge \Gamma \wedge A) \vee \neg A$. Hence by gluing we have $\bigwedge \Gamma \rhd \bigvee \Delta \vee \neg A$.

For (ii), we have $\bigwedge \Gamma \rhd \bigvee \Delta \vee A$. By conjunction application $\bigwedge \Gamma \wedge \neg A \rhd (\bigvee \Delta \vee A) \wedge \neg A$. By the claim we have $A \wedge \neg A \rhd \bot$ therefore by disjunction application

$\bigvee \Delta \vee (A \wedge \neg A) \rhd \bigvee \Delta$. But, it is clear that $(\bigvee \Delta \vee A) \wedge \neg A \geq \bigvee \Delta \vee (A \wedge \neg A)$. Hence by gluing, $\bigwedge \Gamma \wedge \neg A \rhd \bigvee \Delta$.

Now, we will prove both $(*)$ and $(**)$ for the class Σ_{k+1}. For the other two cases for Π_{k+1}, we will use the following duality argument: Note that using negation on all the elements of a $(\Sigma_{k+1}, \mathcal{B}, \sigma)$-flow from C to D provides a $(\Pi_{k+1}, \mathcal{B}, \pi)$-flow from $\neg D$ to $\neg C$. Therefore, the Π_{k+1} case of $(*)$ is provable from the Σ_{k+1} case of $(**)$ and the Π_{k+1} case of $(**)$ is provable from the Σ_{k+1} case of $(*)$.

Assume $\Phi_{k+1} = \Sigma_{k+1}$. For $(*)$, notice that

$$\exists i \leq 1 \, [(i = 1 \to A) \wedge (i = 0 \to \neg A)] \geq_\sigma A \vee \neg A$$

it is enough to witness A and $\neg A$ in both sides with themselves. But since

$$\rhd_\sigma^{(\Sigma_{k+1}, \mathcal{B})} \exists i \leq 1 \, [(i = 1 \to A) \wedge (i = 0 \to \neg A)]$$

by propositional rules and gluing we can deduce $\rhd_\sigma^{(\Sigma_{k+1}, \mathcal{B})} A \vee \neg A$.

For $(**)$, use induction on the complexity of A. If A is quantifier-free, then there is nothing to prove. If $A = B \wedge C$, by IH, $B \wedge \neg B \rhd_\sigma^{(\Sigma_{k+1}, \mathcal{B})}$ and $C \wedge \neg C \rhd_\sigma^{(\Sigma_{k+1}, \mathcal{B})}$ since

$$(B \wedge C) \wedge \neg (B \wedge C) \geq_\sigma (B \wedge \neg B) \vee (C \wedge \neg C)$$

witnessing any quantifier by itself, using gluing we will have

$$(B \wedge C) \wedge \neg (B \wedge C) \rhd_\sigma^{(\Sigma_{k+1}, \mathcal{B})}$$

The case for the disjunction is similar.

If A begins with a universal quantifier, by Lemma 2.4.12, there exists A' such that $A \equiv_\pi A' = \forall \vec{z} \leq \vec{t} \, B(\vec{z}) \in \Pi_k \cup \Sigma_k$ where $\forall \vec{z} \leq \vec{t}$ is the whole left-most block of bounded universal quantifiers and $B \in \Sigma_{k-1}$. Then by the above considerations on duality, since we have

$$\rhd_\sigma^{(\Sigma_k, \mathcal{B})} B(\vec{w}) \vee \neg B(\vec{w})$$

hence

$$B(\vec{w}) \wedge \neg B(\vec{w}) \rhd_\pi^{(\Pi_k, \mathcal{B})}$$

Now by Lemma 2.4.9 we have

$$\exists \vec{w} \leq \vec{t} \forall \vec{z} \leq \vec{t} \, [B(\vec{w}) \wedge \neg B(\vec{w})] \rhd_\sigma^{(\Sigma_{k+1}, \mathcal{B})}$$

Now note that

$$\exists \vec{w} \leq \vec{t} \neg B(\vec{w}) \wedge \forall \vec{z} \leq \vec{t} B(\vec{z}) \geq_\sigma \exists \vec{w} \leq \vec{t} \forall \vec{z} \leq \vec{t} [B(\vec{w}) \wedge \neg B(\vec{w})]$$

because we can witness $\vec{w}$ by itself and $\vec{z}$ by $\vec{w}$. The main point here is that σ-prenex form of $\neg B(\vec{w})$ do not begin with an existential quantifier and hence

after reading the first block of existential quantifiers, the formula $\neg B(\vec{w})$ remains intact. Therefore,

$$\exists \vec{w} \leq \vec{t} \neg B(\vec{w}) \wedge \forall \vec{z} \leq \vec{t} B(\vec{z}) \rhd_{\sigma}^{(\Sigma_{k+1}, \mathcal{B})}$$

hence $A' \wedge \neg A' \rhd_{\sigma}^{(\Sigma_{k+1}, \mathcal{B})}$. Finally, since A begins with at least one universal quantifier and $A \equiv_{\pi} A'$ we have $A \equiv_{\sigma} A'$. On the other hand, $\neg A \equiv_{\sigma} \neg A'$ and hence $A \wedge \neg A \equiv_{\sigma} A' \wedge \neg A'$ which completes the proof.

The case for the existential quantifier is similar. □

In the following lemma, we will show that it is possible to simulate the contraction rule by deterministic reductions in the cost of extending one reduction to a sequence of them, i.e., a flow.

Lemma 2.4.14. *(Structural rules)*

(*i*) *If* $\Gamma, A, B, \Sigma \rhd \Delta$ *then* $\Gamma, B, A, \Sigma \rhd \Delta$.

(*ii*) *If* $\Gamma \rhd \Delta, A, B, \Sigma$ *then* $\Gamma \rhd \Delta, B, A, \Sigma$.

(*iv*) *If* $\Gamma \rhd \Delta$ *then* $\Gamma, A \rhd \Delta$.

(*v*) *If* $\Gamma \rhd \Delta$ *then* $\Gamma \rhd \Delta, A$.

(*iii*) *If* $\Gamma, A, A \rhd \Delta$ *then* $\Gamma, A \rhd \Delta$.

(*vi*) *If* $\Gamma \rhd \Delta, A, A$ *then* $\Gamma \rhd \Delta, A$.

Proof. The weakening and the exchange cases are trivial. For the contraction case notice that in the presence of conjunction and disjunction applications and also the gluing rule, it is enough to prove the following claim:

Claim. For any $\alpha \in \{\pi, \sigma\}$, if $A \in \Phi$, then:

(*i*) $A(\vec{x}) \vee A(\vec{x}) \rhd_{\alpha}^{\Phi} A(\vec{x})$.

(*ii*) $A(\vec{x}) \rhd_{\alpha}^{\Phi} A(\vec{x}) \wedge A(\vec{x})$.

For (i), use induction on the complexity of A. If A is quantifier-free, then there is nothing to prove, because $A \vee A \equiv_{\alpha}^{\Phi} A \equiv_{\alpha}^{\Phi} A \wedge A$.

If $A = B \wedge C$, then by IH, $B \vee B \rhd^{\Phi} B$ and $C \vee C \rhd^{\Phi} C$. But $(B \wedge C) \vee (B \wedge C) \geq (B \vee B) \wedge (C \vee C)$ because it is just enough to witness any quantifier with itself. Hence, by gluing and conjunction application, $(B \wedge C) \vee (B \wedge C) \rhd^{\Phi} B \wedge C$. The case for disjunction is easy.

Now assume $A = \forall z \leq t(\vec{x})\ B(\vec{x}, z)$. If Φ is the class Π_k, by IH we have $B(\vec{x}, z) \vee B(\vec{x}, z) \rhd_{\pi}^{\Pi_k} B(\vec{x}, z)$ and if Φ is Σ_k, since $\forall z \leq t(\vec{x}) B(\vec{x}, z) \in \Phi$, then it actually lives in the lower class Π_{k-1}, which again by IH means $B(\vec{x}, z) \vee B(\vec{x}, z) \rhd_{\pi}^{\Pi_{k-1}} B(\vec{x}, z)$. Hence, in either case

$$B(\vec{x}, z) \vee B(\vec{x}, z) \rhd_{\pi}^{\Pi_k} B(\vec{x}, z)$$

By Lemma 2.4.9 we have

$$\forall z \leq t(\vec{x})\ [B(\vec{x}, z) \vee B(\vec{x}, z)] \rhd_{\alpha}^{\Phi} \forall z \leq t(\vec{x})\ B(\vec{x}, z).$$

for any $\alpha \in \{\sigma, \pi\}$. But $\forall z \leq t(\vec{x})\ [B(\vec{x}, z) \vee B(\vec{x}, z)]$ is α-reducible to

$$\forall u \leq t(\vec{x})\ B(\vec{x}, u) \vee \forall v \leq t(\vec{x}) B(\vec{x}, v)$$

using the variable z as the witness for both of u and v, hence the claim follows from gluing.

For the existential case, w.l.o.g we can assume $\Phi = \Sigma_k$ for some k. The reason is that if $\Phi = \Pi_k$, then since A begins with an existential quantifier, $A \in \Sigma_{k-1}$ and hence we can work with Σ_{k-1}. Therefore, we assume $\Phi = \Sigma_k$ for some k. First note that by the Lemma 2.4.12, there exists $A' = \exists \vec{z} \leq \vec{t}(\vec{x})\ B(\vec{x}, \vec{z})$ such that $A \equiv_{\sigma} A'$. But since both of the formulas A and A' begin with an existential quantifier, $A \equiv_{\pi} A'$. Therefore, it is enough to prove the claim for A'. Note that by this assumption we can assume that the σ-prenex form of B is quantifier-free or begins with universal quantifiers and hence $B \in \Pi_{k-1}$. Then by the Lemma 2.4.13, we have $B(\vec{u}) \wedge \neg B(\vec{u}) \rhd_{\sigma}^{\Sigma_k} \bot$ and $B(\vec{v}) \wedge \neg B(\vec{v}) \rhd_{\sigma}^{\Sigma_k} \bot$ and then by the propositional rules

$$(B(\vec{u}) \vee B(\vec{v})) \wedge \neg B(\vec{u}) \wedge \neg B(\vec{v}) \rhd_{\sigma}^{\Sigma_k} \bot \quad (*)$$

Assume the length of this flow is s. Then, there is a $(\Sigma_k, \mathcal{B}, \sigma)$-flow from

$$(i \leq 1 \wedge j \leq 1) \wedge [B(\vec{u}) \vee B(\vec{v})] \wedge (\chi_B(\vec{u}) = i) \wedge (\chi_B(\vec{v}) = j)$$

to

$$(\chi_B(\vec{u}) = i) \wedge (\chi_B(\vec{v}) = j)] \wedge (i = 1 \vee j = 1)$$

with the length s where $\chi_B(\vec{u}) = i$ means $(i = 1 \rightarrow B(\vec{u})) \wedge (i = 0 \rightarrow \neg B(\vec{u}))$. It is enough to use the formula $G(w, i, j, \vec{u}, \vec{v})$ to fill in between, where G is defined by the following scheme: If $i > 1$ or $j > 1$ then use $\bot$. If $i = j = 1$, then use $G(w, i, j, \vec{u}, \vec{v}) = B(\vec{u}) \wedge B(\vec{v})$. If $i = 1$ and $j = 0$ use $G(w, i, j, \vec{u}, \vec{v}) = B(u) \wedge \neg B(v)$. If $i = 0$ and $j = 1$ use $G(w, i, j, \vec{u}, \vec{v}) = \neg B(u) \wedge B(v)$. And finally if $i = j = 0$, use the flow from $(*)$. Moreover, in the first three cases, use identity reductions, ignoring the $B(\vec{u}) \vee B(\vec{v})$.

Using the Lemma 2.4.9, for any $\alpha \in \{\sigma, \pi\}$ we have a $(\Sigma_k, \mathcal{B}, \alpha)$-flow from

$$\exists \vec{u}, \vec{v} \leq \vec{t}\, \exists i, j \leq 1\ [(i \leq 1 \wedge j \leq 1) \wedge [B(\vec{u}) \vee B(\vec{v})] \wedge (\chi_B(\vec{u}) = i) \wedge (\chi_B(\vec{v}) = j)]$$

to

$$\exists \vec{u}, \vec{v} \leq \vec{t}\, \exists i, j \leq 1\ [(\chi_B(\vec{u}) = i) \wedge (\chi_B(\vec{v}) = j)] \wedge (i = 1 \vee j = 1)]$$

Since the first element of the flow is α-equivalent to

$$\exists \vec{u}, \vec{v} \leq \vec{t}\, [[B(\vec{u}) \vee B(\vec{v})] \wedge \exists i, j \leq 1\ [(\chi_B(\vec{u}) = i) \wedge (\chi_B(\vec{v}) = j)]]$$

for any $\alpha \in \{\sigma, \pi\}$ we will have $(\Sigma_k, \mathcal{B}, \alpha)$-flow from

$$\exists \vec{u}, \vec{v} \leq \vec{t}\, [[B(\vec{u}) \vee B(\vec{v})] \wedge \exists i, j \leq 1\ [(\chi_B(\vec{u}) = i) \wedge (\chi_B(\vec{v}) = j)]]$$

to

$$\exists \vec{u}, \vec{v} \leq \vec{t}\, \exists i, j \leq 1\ [(\chi_B(\vec{u}) = i) \wedge (\chi_B(\vec{v}) = j)] \wedge (i = 1 \vee j = 1)]$$

On the other hand, by the Lemma 2.4.11 and the Lemma 2.4.9, for any $\alpha \in \{\sigma, \pi\}$ we know that there is a $(\Sigma_k, \mathcal{B}, \alpha)$-flow from

$$\exists \vec{u}, \vec{v} \leq \vec{t}\, B(\vec{u}) \vee B(\vec{v})$$

to

$$\exists \vec{u}, \vec{v} \leq \vec{t}\, [[B(\vec{u}) \vee B(\vec{v})] \wedge \exists i, j \leq 1\ [(\chi_B(\vec{u}) = i) \wedge (\chi_B(\vec{v}) = j)]]$$

Now, since

$$\exists \vec{u}, \vec{v} \leq \vec{t}\, B(\vec{u}) \vee B(\vec{v})$$

and

$$\exists \vec{u} \leq \vec{t}\, B(\vec{u}) \vee \exists \vec{v} \leq \vec{t}\, B(\vec{v})$$

are α-equivalent, it is enough to show that $\exists \vec{y} \leq \vec{t}(\vec{x})\ B(\vec{x}, \vec{y})$ is α-reducible to

$$\exists \vec{u}, \vec{v} \leq \vec{t}(\vec{x})\ \exists i, j \leq 1\ (i = 1 \vee j = 1) \wedge (\chi_B(u) = i) \wedge (\chi_B(v) = j)$$

It is enough to read i and j and decide between the cases that $i = 1$ or ($i = 0$, $j = 1$). Then if $i = 1$, use $\vec{u}$ to witness $\vec{y}$ and reduce $B(\vec{y})$ to $B(\vec{u})$ in $\chi_B(\vec{u}) = i$ by identity reduction. If ($i = 0$, $j = 1$) then use $\vec{v}$ to witness $\vec{y}$ and reduce $B(\vec{y})$ to $B(\vec{v})$ in $\chi_B(\vec{v}) = j$ by identity reduction.

The case (ii) is the dual of (i) and provable by just using (i) on $\neg A$ and then taking negations. □

Lemma 2.4.15. *(Cut and Induction)*

(i) *If* $\Gamma_0(\vec{x}) \rhd A(\vec{x}), \Delta_0(\vec{x})$ *and* $\Gamma_1(\vec{x}), A(\vec{x}) \rhd \Delta_1(\vec{x})$*, then we have* $\Gamma_0(\vec{x}), \Gamma_1(\vec{x}) \rhd \Delta_0(\vec{x}), \Delta_1(\vec{x})$.

(ii) *If* $\Gamma(\vec{x}), A(y, \vec{x}) \rhd \Delta(\vec{x}), A(y+1, \vec{x})$ *then* $\Gamma(\vec{x}), A(0, \vec{x}) \rhd \Delta(\vec{x}), A(s(\vec{z}, \vec{x}), \vec{x})$.

Proof. For (i), Since $\Gamma_0 \rhd \Delta_0, A$ and $\Gamma_1, A \rhd \Delta_1$ then

$$\bigwedge \Gamma_0 \rhd \bigvee \Delta_0 \vee A$$

and $\bigwedge \Gamma_1 \wedge A \rhd \bigvee \Delta_1$. Apply conjunction with $\bigwedge \Gamma_1$ on the first one and disjunction with $\bigvee \Delta_0$ on the second one to prove $\bigwedge \Gamma_1 \wedge \bigwedge \Gamma_0 \rhd (\bigvee \Delta_0 \vee A) \wedge \bigwedge \Gamma_1$ and $(\bigwedge \Gamma_1 \wedge A) \vee \bigvee \Delta_0 \rhd \bigvee \Delta_1 \vee \bigvee \Delta_0$. Since $(\bigvee \Delta_0 \vee A) \wedge \bigwedge \Gamma_1 \geq (\bigwedge \Gamma_1 \wedge A) \vee \bigvee \Delta_0$, by using gluing we will have $\bigwedge \Gamma_1 \wedge \bigwedge \Gamma_0 \rhd \bigvee \Delta_0 \vee \bigvee \Delta_1$.

For (ii) we reduce the induction case to the strong gluing case. Since

$$\Gamma, A(y, \vec{x}) \rhd \Delta, A(y+1, \vec{x})$$

by definition, $\bigwedge \Gamma \wedge A(y, \vec{x}) \rhd \bigvee \Delta \vee A(y+1, \vec{x})$. Therefore, by the Lemma 2.4.7 we have

$$(\bigwedge \Gamma \wedge A(y, \vec{x})) \vee \bigvee \Delta \rhd \bigvee \Delta \vee A(y+1, \vec{x}) \vee \bigvee \Delta$$

and by contraction for $\bigvee \Delta$ we know

$$\bigvee \Delta \vee A(y+1, \vec{x}) \vee \bigvee \Delta \rhd \bigvee \Delta \vee A(y+1, \vec{x}).$$

Hence,

$$(\bigwedge \Gamma \wedge A(y,\vec{x})) \vee \bigvee \Delta \rhd \bigvee \Delta \vee A(y+1,\vec{x}).$$

Then by conjunction introduction and the fact that $(\bigwedge \Gamma \wedge A(y,\vec{x})) \vee \bigvee \Delta) \rhd \bigwedge \Gamma \vee \bigvee \Delta$,

$$((\bigwedge \Gamma \wedge A(y,\vec{x})) \vee \bigvee \Delta), (\bigwedge \Gamma \wedge A(y,\vec{x})) \vee \bigvee \Delta) \rhd (\bigvee \Delta \vee A(y+1,\vec{x})) \wedge (\bigwedge \Gamma \vee \bigvee \Delta)$$

By using the propositional, structural and the cut rule, it is easy to prove

$$(\phi \vee \psi) \wedge (\sigma \vee \psi) \rhd (\phi \wedge \sigma) \vee \psi.$$

Hence, by using the contraction we have

$$(\bigwedge \Gamma \wedge A(y,\vec{x})) \vee \bigvee \Delta \rhd (\bigwedge \Gamma \wedge A(y+1,\vec{x})) \vee \bigvee \Delta.$$

Now by strong gluing we have

$$(\bigwedge \Gamma \wedge A(0,\vec{x})) \vee \bigvee \Delta \rhd (\bigwedge \Gamma \wedge A(s(\vec{z},\vec{x}),\vec{x})) \vee \bigvee \Delta.$$

But since $\Gamma \wedge A(0,\vec{x}) \rhd (\bigwedge \Gamma \wedge A(0,\vec{x})) \vee \bigvee \Delta$ and

$$(\bigwedge \Gamma \wedge A(s(\vec{x}),\vec{x})) \vee \bigvee \Delta \geq \bigvee \Delta \vee A(s(\vec{z},\vec{x}),\vec{x}),$$

we have

$$\Gamma(\vec{x}), A(0,\vec{x}) \rhd \Delta(\vec{x}), A(s(\vec{z},\vec{x}),\vec{x}).$$

□

Lemma 2.4.16. *(Implication Rules) If $A \to B \in \Phi$:*

(i) If $\Gamma_0 \rhd^\Phi \Delta_0, A$ and $\Gamma_1, B \rhd^\Phi \Delta_1$ then $\Gamma_0, \Gamma_1, A \to B \rhd^\Phi \Delta_0, \Delta_1$.

(ii) If $\Gamma, A \rhd^\Phi \Delta, B$ then $\Gamma \rhd^\Phi \Delta, A \to B$.

Proof. Notice that if $A \to B \in \Phi$ then $A \to B$ is quantifier-free and hence $\neg A, B \in \Phi$. Therefore, by definition, it is easy to see that $A \to B \equiv \neg A \vee B$. Therefore:

For (i) since $\Gamma_0 \rhd \Delta_0, A$ by the Lemma 2.4.13 we have $\Gamma_0, \neg A \rhd \Delta_0$. On the other hand, we have $\Gamma_1, B \rhd \Delta_1$. Therefore, by the Lemma 2.4.10 we have $\Gamma_0, \Gamma_1, \neg A \vee B \rhd \Delta_0, \Delta_1$. Since $A \to B \rhd \neg A \vee B$, by using cut we have $\Gamma_0, \Gamma_1, A \to B \rhd \Delta_0, \Delta_1$.

For (ii), since we have $\Gamma, A \rhd \Delta, B$ then by the Lemma 2.4.13 we will have $\Gamma, \rhd\Delta, \neg A, B$. Hence by the Lemma 2.4.10 we have $\Gamma, \rhd\Delta, (\neg A \vee B), (\neg A \vee B)$. By contraction, $\Gamma, \rhd\Delta, (\neg A \vee B)$. Since $\neg A \vee B \rhd A \to B$, by cut $\Gamma, \rhd\Delta, A \to B$. □

The following theorem is the main theorem of the theory of flows for bounded theories of arithmetic:

Theorem 2.4.17. *(Soundness) If $\Gamma(\vec{x}) \cup \Delta(\vec{x}) \subseteq \Phi$, $\mathfrak{B}(\Phi, \mathcal{A}) \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$ and $\mathcal{A} \subseteq \mathcal{B}$ has a characteristic function for any quantifier-free formula then $\Gamma \rhd_\phi^{(\Phi,\mathcal{B})} \Delta$.*

Proof. We assume Φ is a π-type class. The other case is similar. To prove the lemma we use induction on the length of the free-cut free proof of $\Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$.

1. (Axioms). If $\Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$ is a logical axiom then the claim is trivial. If it is a non-logical axiom then the claim will be also trivial because all non-logical axioms are quantifier-free and provable in $\mathcal{B}$. Therefore there is nothing to prove.

2. (Structural Rules). It is proved in the Lemma 2.4.14.

3. (Cut). It is proved by Lemma 2.4.15.

4. (Propositional). The conjunction and disjunction cases are proved in the Lemma 2.4.10. The implication and negation cases are proved in the Lemma 2.4.16.

5. (Bounded Universal Quantifier Rules, Right). If $\Gamma(\vec{x}) \Rightarrow \Delta(\vec{x}), \forall z \leq p(\vec{x})B(\vec{x}, z)$ is proved by the $\forall^{\leq}R$ rule by $\Gamma(\vec{x}), b \leq p(\vec{x}) \Rightarrow \Delta(\vec{x}), B(\vec{x}, b)$, then by IH, $\Gamma(\vec{x}), b \leq p(\vec{x}) \rhd_{\pi}^{\Pi_k} \Delta(\vec{x}), B(\vec{x}, b)$. By the Lemma 2.4.9, we have a $(\Pi_k, \mathcal{B}, \pi)$-flow from $\forall b \leq p(\vec{x})(b \leq p(\vec{x}) \wedge \bigwedge \Gamma)$ to $\forall b \leq p(\vec{x})[B(\vec{x}, b) \vee \bigvee \Delta]$. Since Γ does not have a free b, it is easy to see that $\bigwedge \Gamma \geq_{\pi} \forall b \leq p(\vec{x})(b \leq p(\vec{x}) \wedge \bigwedge \Gamma)$. Hence it is enough to add $\bigwedge \Gamma$ to the beginning of the flow. Do the same for the right side to reach $\forall b \leq p(\vec{x})B(\vec{x}, b) \vee \bigvee \Delta$. Finally note that changing the name of a bounded variable does not change the nature of deterministic flows which complete the proof.

6. (Bounded Universal Quantifier Rules, Left). Suppose

$$\Gamma(\vec{x}), s(\vec{x}) \leq p(\vec{x}), \forall z \leq p(\vec{x})B(\vec{x}, z) \Rightarrow \Delta(\vec{x})$$

is proved by the $\forall^{\leq}L$ rule by $\Gamma(\vec{x}), B(\vec{x}, s(\vec{x})) \Rightarrow \Delta(\vec{x})$. Then by IH,

$$\Gamma(\vec{x}), B(\vec{x}, s(\vec{x})) \rhd_{\pi}^{\Pi_k} \Delta(\vec{x})$$

But by witnessing z by s and the rest by themselves, we have

$$\bigwedge \Gamma(\vec{x}) \wedge s(\vec{x}) \leq p(\vec{x}) \wedge \forall z \leq p(\vec{x})B(\vec{x}, z) \geq_{\pi} \bigwedge \Gamma(\vec{x}) \wedge B(\vec{x}, s(\vec{x}))$$

hence by gluing

$$\Gamma(\vec{x}), s(\vec{x}) \leq p(\vec{x}), \forall z \leq p(\vec{x})B(\vec{x}, z) \geq_{\pi} \Delta(\vec{x}).$$

7. (Bounded Existential Quantifier Rules, Right). If $\Gamma(\vec{x}), s(\vec{x}) \leq p(\vec{x}) \Rightarrow \Delta(\vec{x}), \exists z \leq p(\vec{x})B(\vec{x}, z)$ is proved by the $\exists^{\leq}R$ rule by $\Gamma(\vec{x}) \Rightarrow \Delta(\vec{x}), B(\vec{x}, s(\vec{x}))$ then by IH

$$\Gamma(\vec{x}) \rhd_{\pi}^{\Pi_k} \Delta(\vec{x}), B(\vec{x}, s(\vec{x})).$$

Since $\exists z \leq p(\vec{x})B(\vec{x}, z) \in \Pi_k$, it is also in Σ_{k-1}. Therefore, by Lemma 2.4.13, $\Gamma(\vec{x}), \neg B(\vec{x}, s(\vec{x})) \rhd_{\pi}^{\Pi_k} \Delta(\vec{x})$. By 6, $\Gamma(\vec{x}), s(\vec{x}) \leq p(\vec{x}), \forall z \leq p(\vec{x}) \neg B(\vec{x}, z) \rhd_{\pi}^{\Pi_k} \Delta(\vec{x})$ and again by the Lemma 2.4.13 we will have

$$\Gamma(\vec{x}), s(\vec{x}) \leq p(\vec{x}) \rhd_{\pi}^{\Pi_k} \Delta(\vec{x}), \exists z \leq p(\vec{x})B(\vec{x}, z).$$

8. (Bounded Existential Quantifier Rules, Left). If $\Gamma, \exists y \leq p(\vec{x}) B(\vec{x}, y) \Rightarrow \Delta$ is proved by the $\exists^{\leq} L$ rule by $\Gamma, b \leq p(\vec{x}), B(\vec{x}, b) \Rightarrow \Delta$, by IH we have $\Gamma, b \leq p(\vec{x}), B(\vec{x}, b) \rhd_{\pi}^{\Pi_k} \Delta$ then since $\exists b \leq p(\vec{x}) B(\vec{x}, b) \in \Pi_k$, it is also in Σ_{k-1}. Therefore, by the Lemma 2.4.13

$$\Gamma, b \leq p(\vec{x}) \rhd_{\pi}^{\Pi_k} \Delta, \neg B(\vec{x}, b)$$

by 5, we have

$$\Gamma \rhd_{\pi}^{\Pi_k} \Delta, \forall y \leq p(\vec{x}) \neg B(\vec{x}, y)$$

Finally again by Lemma 2.4.13 we have

$$\Gamma, \exists y \leq p(\vec{x}) B(\vec{x}, y) \rhd_{\pi}^{\Pi_k} \Delta.$$

9. (Induction). It is proved in Lemma 2.4.15. □

We also have the following completeness theorem:

Theorem 2.4.18. *(Completeness) If* $\Gamma(\vec{x}) \rhd_{\phi}^{(\Phi_k, \mathcal{B})} \Delta(\vec{x})$ *and* $\mathcal{B} \subseteq \mathfrak{B}(\Phi_k, \mathcal{A})$*, then* $\mathfrak{B}(\Phi_k, \mathcal{A}) \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$.

Proof. If $\Gamma(\vec{x}) \rhd_{\phi}^{(\Phi_k, \mathcal{B})} \Delta(\vec{x})$, then by the fact that the existence of deterministic reductions implies provability, there exist a term $t(\vec{x})$, and a formula $H(u, \vec{x}) \in \Phi_k$ such that we have the following:

(*i*) $\mathcal{B} \vdash H(0, \vec{x}) \leftrightarrow \bigwedge \Gamma(\vec{x})$,

(*ii*) $\mathcal{B} \vdash H(t(x), \vec{x}) \leftrightarrow \bigvee \Delta(\vec{x})$,

(*iii*) $\mathcal{B} \vdash \forall u \leq t(\vec{x})\ H(u, \vec{x}) \rightarrow H(u+1, \vec{x})$.

Since $\mathcal{B} \subseteq \mathfrak{B}(\Phi_k, \mathcal{A})$, we have

$$\mathfrak{B}(\Phi_k, \mathcal{A}) \vdash \forall u \leq t(\vec{x})\ H(u, \vec{x}) \rightarrow H(u+1, \vec{x}).$$

Since $H(u, \vec{x}) \in \Phi_k$ by induction we have,

$$\mathfrak{B}(\Phi_k, \mathcal{A}) \vdash H(0, \vec{x}) \rightarrow H(t(\vec{x}), \vec{x}).$$

On the other hand, we have $\mathcal{B} \vdash H(0, \vec{x}) \leftrightarrow \bigwedge \Gamma(\vec{x})$ and $\mathcal{B} \vdash H(t(\vec{x}), \vec{x}) \leftrightarrow \bigvee \Delta(\vec{x})$. Therefore, $\mathfrak{B}(\Phi_k, \mathcal{A}) \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$. □

2.4.3 Applications

In this subsection we will use the soundness and completeness theorems that we have proved in the previous subsection to extract the computational content of the low complexity statements of some concrete weak bounded theories such as Buss' hierarchy of bounded theories of arithmetic and some strong theories such as $I\Delta_0(\exp)$, PRA and PA + TI(α).

For the first application, consider the theories $IU_k = \mathfrak{B}(\Pi_k(\mathcal{L}_{\mathcal{R}}), \mathcal{R})$ for $k \geq 1$. These theories are the fragments of the theory $I\Delta_0$ corresponding to the computational world of the linear time hierarchy. Moreover, consider the class of all functions constructed from zero, projections and closed under successor, addition, production, subtraction and division and call it R:

Corollary 2.4.19. *Let* $\Gamma(\vec{x})\cup\Delta(\vec{x}) \subseteq U_k$. *Then,* $IU_k \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$ *iff* $\Gamma \rhd_\pi^{(U_k,\mathcal{R})} \Delta$. *The second condition means that there exists a sequence of length* $t \in R$ *of formulas in* U_k *beginning from* $\bigwedge\Gamma$ *ending with* $\bigvee\Delta$ *such that each formula is* $(\pi,\mathcal{R})$*-reducible to its successor using just the functions in* R.

Proof. The only thing that we have to check is the fact that $\mathcal{R}$ has the characteristic functions for any quantifier-free formula in the language $\mathcal{L}_\mathcal{R}$. It has been proved in the Remark 2.2.2. □

The second application, and maybe the more important one, is the case of Buss' hierarchy of bounded arithmetic, in which we assume the language has a symbol for any PV function and we denote the class of all strict Σ_k^b and Π_k^b formulas with $\hat{\Sigma}_k^b$ and $\hat{\Pi}_k^b$.

Corollary 2.4.20. *Let* $\Gamma(\vec{x}) \cup \Delta(\vec{x}) \subseteq \hat{\Pi}_k^b(\#_n)$. *Then,* $T_n^k \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$ *iff* $\Gamma \rhd_\pi^{(\hat{\Pi}_k^b(\#_n),\mathrm{BASIC}_n(\mathrm{PV}))} \Delta$, *where* $\mathrm{BASIC}_n(\mathrm{PV}))$ *is the theory* BASIC_n *plus all the defining axioms of* PV. *Specifically, for* $n = 2$, $T_2^k \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$ *iff* $\Gamma \rhd_\pi^{(\hat{\Pi}_k^b,\mathrm{PV})} \Delta$. *The second condition in the latter case means the existence of a uniform sequence of length* $2^{p(|\vec{x}|)}$ *of formulas in* Π_k^b *starting with* $\bigwedge\Gamma$ *and ending in* $\bigvee\Delta$ *such that each formula is* (π,PV)*-reducible to its successor, using just the polynomial time computable functions.*

Proof. Observe that in the presence of all PV functions, any formula in $\hat{\Pi}_k^b(\#_n)$ is equivalent to a formula in Π_k. Therefore, since T_n^k is axiomatizable by $\hat{\Pi}_k^b(\#_n)$-induction, it is also axiomatizable by Π_k-induction. □

And also we can apply the soundness theorem on stronger theories with full exponentiation like $I\Delta_0(\exp)$ and PRA. Consider the theory $\mathcal{R}$ augmented with a function symbol for exponentiation with the usual recursive definition and denote it by $\mathcal{R}(\exp)$ and also denote the union of $\mathcal{R}$ and the induction-free part of PRA by PRA^-. Then:

Corollary 2.4.21. *Let* $\Gamma(\vec{x}) \cup \Delta(\vec{x}) \subseteq \Pi_k$. *Then:*

(*i*) $I\Delta_0(\exp) \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$ *iff* $\Gamma \rhd_\pi^{(\Pi_k,\mathcal{R}(\exp))} \Delta$.

(*ii*) $\mathrm{PRA} \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$ *iff* $\Gamma \rhd_\pi^{(\Pi_k,\mathrm{PRA}^-)} \Delta$.

Proof. The only point to mention is that both of the theories $I\Delta_0(\exp)$ and PRA are axiomatizable by $I\Pi_k$ for any k. Hence we can apply the theory of deterministic flows here. □

We can also use the theory of flows to extract the computational content of low complexity sentences of the very strong theories of arithmetic such as PA and $\mathrm{PA} + \mathrm{TI}(\alpha)$. But this is not what we can implement in a very direct way. The reason is that our method is tailored for bounded theories while these theories are unbounded. Hence, to use our theory, we have to find a way to transfer low complexity statements from these theories to some corresponding bounded theories. This is what the continuous cut elimination method makes possible in its very elegant enterprise. It transfers all Π_2^0 consequences of a strong theory T to some quantifier-free extensions of PRA and then makes it possible to apply the flow decomposition technique. To explain how it works, we need some definitions:

Definition 2.4.22. *Let T be a theory of arithmetic. We say that α is a Π^0_2-proof theoretical ordinal of T when $\prec$ is the primitive recursive representation of the order on α and $T \equiv_{\Pi^0_2} \mathrm{PRA} + \bigcup_{\beta \prec \alpha} \mathrm{TI}(\prec_\beta)$ where $\mathrm{TI}(\prec_\beta)$ means full transfinite induction up to the ordinal β.*

Convention. From now on wherever we have a proof theoretic ordinal, we always assume that it is closed under the operation $\beta \mapsto \omega^\beta$.

Definition 2.4.23. *Let $\prec$ be a quantifier-free formula in the language of* PRA. *By theory* $\mathrm{PRA} + \bigcup_{\beta \prec \alpha} \mathrm{PRWO}(\prec_\beta)$ *we mean* PRA *plus the axiom schema*

$$\mathrm{PRWO}(\prec_\beta) : \forall \vec{x} \exists y \; f(\vec{x}, y+1) \not\prec_\beta f(\vec{x}, y)$$

for any function symbol f.

The following theorem uses continuous cut elimination technique to reduce transfinite induction to PRWO.

Theorem 2.4.24. *[11] Let T be a theory of arithmetic and α its Π^0_2-proof theoretical ordinal. Then*

$$T \equiv_{\Pi^0_2} \mathrm{PRA} + \bigcup_{\beta \prec \alpha} \mathrm{PRWO}(\prec_\beta)$$

The following theory is the skolemization of $\mathrm{PRA} + \bigcup_{\beta \prec \alpha} \mathrm{PRWO}(\prec_\beta)$:

Definition 2.4.25. *The language of the theory* $\mathrm{PRA}_\prec$ *consists of the language of* PRA *plus the scheme which says that for any* PRA*-function symbol $f(\vec{x}, y)$ and any $\beta \prec \alpha$, there exists a function symbol $[\mu_\beta y.f](\vec{x})$. Then* $\mathrm{BASIC}_\prec$ *is the theory axiomatized by the axioms of* PRA *plus the theory $\mathcal{R}$ and the following definitional equations: $f(\vec{x}, 1+[\mu_\beta y.f](\vec{x})) \not\prec_\beta f(\vec{x}, [\mu_\beta y.f](\vec{x}))$ and $z < [\mu_\beta y.f](\vec{x}) \to f(\vec{x}, z+1) \prec_\beta f(\vec{x}, z)$. Finally,* $\mathrm{PRA}_\prec$ *is* $\mathrm{BASIC}_\prec$ *plus the usual quantifier-free induction.*

Combining all of these steps together we can reduce a theory T to a bounded arithmetical theory $\mathrm{PRA}_\prec$.

Corollary 2.4.26. *Let T be a theory of arithmetic and α its Π^0_2-proof theoretical ordinal. Then $T \equiv_{\Pi^0_2} \mathrm{PRA}_\prec$.*

Now we are ready to have the following corollary:

Corollary 2.4.27. *Let $\Gamma(\vec{x}) \cup \Delta(\vec{x}) \subseteq \Pi_k$, and α_T is the Π^0_2-ordinal of T with the primitive recursive representation $\prec_{\alpha_T}$, then $T \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$ iff*

$$\Gamma(\vec{x}) \rhd_\pi^{(\Pi_k, \mathrm{BASIC}_{\prec_{\alpha_T}})} \Delta(\vec{x}).$$

Proof. Note that the existence of the flow is equivalent to the provability of $\Gamma \Rightarrow \Delta$ in $\mathrm{PRA}_{\prec_{\alpha_T}}$ because $\mathrm{PRA}_{\prec_{\alpha_T}}$ is a bounded theory axiomatizable by the usual induction on formulas in Π_k. On the other hand, we have $\Gamma(\vec{x}) \cup \Delta(\vec{x}) \subseteq \Pi_k$. Hence the sequent is bounded and is in Π^0_2. Therefore, by the definition of Π^0_2-ordinals, we know that $\mathrm{PRA}_{\prec_{\alpha_T}} \vdash \Gamma \Rightarrow \Delta$ iff $T \vdash \Gamma \Rightarrow \Delta$, which completes the proof. □

Corollary 2.4.28. *Let $\Gamma(\vec{x}) \cup \Delta(\vec{x}) \subseteq \Pi_k$, and $\epsilon(\alpha)$ be the least ϵ number after α with a primitive recursive representation. Then* $\mathrm{PA} + \mathrm{TI}(\alpha) \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$ *iff*

$$\Gamma(\vec{x}) \rhd_\pi^{(\Pi_k, \mathrm{BASIC}_{\prec \epsilon(\alpha)})} \Delta(\vec{x}).$$

So far, we have used the theory of deterministic flows to decompose first order proofs of bounded theories. In the following we will introduce two different kinds of characterizations and we will use them to reprove some recent results for some specific classes of formulas. The types that we want to use are generalizations of some recent characterizations of some low complexity statements in Buss' hierarchy of bounded arithmetic by Game induction principles [19], [21] and some kind of PLS problems [6].

First let us generalize our game interpretation of the Remark 2.3.4 to interpret any formula of the form

$$A = \forall \vec{y_1} \leq \vec{p_1}(\vec{x}) \exists \vec{z_1} \leq \vec{q_1}(\vec{x}) \forall \vec{y_2} \leq \vec{p_2}(\vec{x}) \ldots G_A(\vec{x}, \vec{y_1}, \vec{z_1}, \vec{y_2}, \vec{z_2}, \ldots)$$

as a k-turn game $\mathcal{G}_A$ in which the players can have some but fixed predefined number of simultaneous moves. More precisely, in the game $\mathcal{G}_A$, the first player begins by choosing the moves $\vec{y_1} \leq \vec{p_1}(\vec{x})$ altogether, then the second player chooses the moves $\vec{z_1} \leq \vec{q_1}(\vec{x})$ and they continue alternately. Again if $G_A(\vec{x}, \vec{y_1}, \vec{z_1}, \vec{y_2}, \vec{z_2}, \ldots)$ becomes true the second player wins and otherwise the first player is the winner. Note that in this multi-move version, we still have the equivalence between the truth of A and the existence of the winning strategy for the second player. What we want to add to this fact is its explicit version which states that any deterministic reduction from A to $\top$ is nothing but an explicit winning strategy for the second player in the game $\mathcal{G}_A$.

Definition 2.4.29. *Let $\mathcal{L} \supseteq \mathcal{L}_\mathcal{R}$ be a language. An instance of the (j,k)-game induction principle, $GI_k^j(\mathcal{L})$, is given by size parameters a and b, a quantifier-free formula $G(u, \vec{v})$ with a fixed partition of the variables $\vec{v}$ into k groups, a sequence of terms V and a uniform sequence W_u of sequences of terms. The instance $GI(G, V, W, a, b)$ states that, interpreting $G(u, \vec{v})$ as a k-turn game on moves $\vec{v}$ in which all moves are bounded by b, the following cannot all be true:*

- *(i) Deciding the winner of the game $G(0, \vec{v})$ depends only on the first j moves,*
- *(ii) The second player has a winning strategy for $G(0, \vec{v})$ (expressed as a Π_j formula.)*
- *(iii) For $u \leq a \dot{-} 2$, W_u gives a deterministic reduction from $G(u+1, \vec{v})$ to $G(u, \vec{v})$,*
- *(iv) V is an explicit winning strategy for the first player in $G(a \dot{-} 1, \vec{v})$.*

Notation. Let $\mathcal{C}$ and $\mathcal{D}$ be two classes of formulas and $\mathcal{B}$ be a theory. By $\mathcal{C} \equiv_\mathcal{B} \mathcal{D}$ we mean that for any $A \in \mathcal{C}$ there exists $B \in \mathcal{D}$ such that $B \geq_\pi A$ and for any $A \in \mathcal{D}$ there exists $B \in \mathcal{C}$ such that $B \geq_\pi A$.

Theorem 2.4.30. *Let $j \leq k$. Then,*

$$\forall\Sigma_j[\mathfrak{B}(\Pi_k, \mathcal{B})] \equiv_{\mathcal{B}} GI_k^j(\mathcal{L}).$$

Proof. It is clear that $\mathfrak{B}(\Pi_k, \mathcal{B}) \vdash GI_k^j(\mathcal{L})$ and $GI_k^j(\mathcal{L})$ is expressible by a $\forall\Sigma_j$ sentence. For the converse, assume $\mathfrak{B}(\Pi_k, \mathcal{B}) \vdash \forall\vec{x}A(\vec{x})$ where $A \in \Sigma_j$ and $j \leq k$. Then, we know that $\mathfrak{B}(\Pi_k, \mathcal{B}) \vdash \neg A(\vec{x}) \Rightarrow \bot$ and $\neg A \in \Pi_j$. By Corollary 2.4.20, there exist a term $t(\vec{x})$, a formula $H(u, \vec{x}) \in \Pi_k$ and sequences of terms E_0, E_1, I_0, I_1 and $F(u)$ such that the following statements are provable in $\mathcal{B}$:

(*i*) $H(0, \vec{x}) \equiv_\pi^{(E_0,E_1)} \neg A(\vec{x})$.

(*ii*) $H(t(\vec{x}), \vec{x}) \equiv_\pi^{(I_0,I_1)} \bot$.

(*iii*) $\forall u < t(\vec{x})[H(u, \vec{x}) \geq_\pi^{F_u} H(u+1, \vec{x})]$.

First of all, note that we can change the definition of H in the following way:

$$H'(u, \vec{x}) = (u = 0 \to \neg A(\vec{x})) \wedge (u \neq 0 \to H(u \dot{-} 1, \vec{x})).$$

And, it is possible to shift also the reductions to have (*i*) to (*iii*) for H'. Call these reductions E_0', E_1', F_u', I_0' and I_1'. Note that the truth of $H'(0, x)$ depends only on first j blocks of quantifiers when we write it in the Π_k form.

W.l.o.g., we assume that all bounds in $H'(u, \vec{x})$ are the same and depend only on $\vec{x}$. Call this bound $s(\vec{x})$. This is possible because any term is majorizable by a monotone term. Again w.l.o.g we can assume that H' is in the following prenex form:

$$H'(u, \vec{x}) = \forall\vec{z}_1 \leq s\exists\vec{y}_1 \leq s\forall\vec{z}_2 \leq s \ldots G(u, \vec{x}, \vec{z}_1, \vec{y}_1, \vec{z}_2, \ldots)$$

where G is quantifier-free and the number of quantifier groups are k. Define $a = t(\vec{x})$, $b = s(\vec{x})$, $W_u = F_u'$ and $V = I_0'$ and pick G for the game predicate with its natural partition of variables. Therefore, we have an instance of the game induction. Now we want to show that $A(\vec{x})$ is reducible to this game induction provably in $\mathcal{B}$. Since $\mathcal{B} \vdash \forall u < t(\vec{x})[H'(u, \vec{x}) \geq_\pi^{F_u'} H'(u+1, \vec{x})]$ and $\mathcal{B} \vdash H(t(\vec{x}), \vec{x}) \equiv^{(I_0,I_1)} \bot$, the false part of $GI_k^j(\mathcal{L})$ is the part which states "The second player has a winning strategy for $G(0, \vec{v})$." which means that $H'(0, \vec{x})$ is false. Since $H'(0, \vec{x})$ is equivalent with $\neg A(\vec{x})$ provably in $\mathcal{B}$, the reduction of the sentence $A(\vec{x})$ to the game induction principle is proved. □

Using this generalization it is trivial to reprove the case for Buss' hierarchy of bounded arithmetic:

Corollary 2.4.31. *([19], [21]) For all $j \leq k$, $\forall\Sigma_j(T_2^k) \equiv_{\mathrm{PV}} GI_k^j(\mathcal{L}_{\mathrm{PV}})$.*

Now, let us explain the second type of problems, i.e., the generalized local search problems:

Definition 2.4.32. *A formalized $(\Psi, \Lambda, \mathcal{B}, \prec)$-GLS problem consists of the following data:*

(*i*) *A sequence of terms $\vec{N}(\vec{x}, \vec{s}) \in \mathcal{L}_{\mathcal{B}}$ as the local improvements.*

(*ii*) *A term* $c(\vec{x},\vec{s}) \in \mathcal{L}_{\mathcal{B}}$ *as a cost function.*

(*iii*) *A predicate* $F(\vec{x},\vec{s}) \in \Psi$ *which intuitively means that* $\vec{s}$ *is a feasible solution for the input* $\vec{x}$*.*

(*iv*) *An initial sequence of terms* $\vec{i}(\vec{x}) \in \mathcal{L}_{\mathcal{B}}$*.*

(*v*) *A goal predicate* $G(\vec{x},\vec{s'}) \in \Lambda$*.*

(*vi*) *A quantifier-free predicate* $\prec \in \mathcal{L}_{\mathcal{B}}$ *as a well-ordering.*

(*vii*) *A sequence of bounding terms* $\vec{t}(\vec{x})$*.*

(*viii*) *A projection function* I*.*

such that $\mathcal{B}$ *proves that* $\prec$ *is a total order and*

$$\mathcal{B} \vdash \forall \vec{x}\ F(\vec{x},\vec{i}(\vec{x}))$$

$$\mathcal{B} \vdash \forall \vec{x}\vec{s}\ (F(\vec{x},\vec{s}) \rightarrow F(\vec{x},\vec{N}(\vec{x},\vec{s})))$$

$$\mathcal{B} \vdash \forall \vec{x}\vec{s}\ (\vec{N}(\vec{x},\vec{s}) = \vec{s} \vee c(\vec{x},\vec{N}(\vec{x},\vec{s})) \prec c(\vec{x},\vec{s}))$$

$$\mathcal{B} \vdash \forall \vec{x}\vec{s}\ ((\vec{N}(\vec{x},\vec{s}) = \vec{s} \wedge F(\vec{x},\vec{s})) \rightarrow G(\vec{x},I(\vec{s})))$$

$$\mathcal{B} \vdash \forall \vec{x}\vec{s'}\ (G(\vec{x},\vec{s'}) \rightarrow \vec{s'} \leq \vec{t}(\vec{x}))$$

By the computational problem associated to a GLS *problem, we mean finding* $\vec{s'} \leq \vec{t}(\vec{x})$ *such that* $G(\vec{x},\vec{s'})$*.*

If there is also a sequence of terms $\vec{b}(\vec{x})$ *such that*

$$\mathcal{B} \vdash \forall \vec{x}\vec{s}\ (F(\vec{x},\vec{s}) \rightarrow \vec{s} \leq \vec{b}(\vec{x}))$$

The GLS*-problem is called bounded and their class is denoted by* $\mathrm{BGLS}(\Psi,\Lambda,\mathcal{B},\prec)$*. Moreover, if* $\mathcal{L}_{\mathrm{PV}} \subseteq \mathcal{L}_{\mathcal{B}}$ *and* $\vec{t}(\vec{x}) = 2^{\vec{p}(|x|)}$ *for some polynomials* $\vec{p}$ *we denote the class by* $\mathrm{PLS}(\Psi,\Lambda,\prec,\mathcal{B})$ *and if we have also the conditions that* F *is quantifier-free in the language of* $\mathcal{B}$ *and* G *is quantifier-free in the language of* PV*, we denote the class by* $\mathrm{PLS}(\prec,\mathcal{B})$*. Finally if we also add* $\mathcal{B} = \mathrm{PV}$*, then we write* $\mathrm{PLS}(\prec)$ *for the class of these* GLS*-problems.*

Theorem 2.4.33. (*i*) *For any* $\mathrm{BGLS}(\Pi_k,\Lambda,\mathcal{B},\leq)$*-problem we have:*

$$\mathfrak{B}(\Pi_{k+1},\mathcal{B}) \vdash \forall \vec{x} \exists \vec{s'} G(\vec{x},\vec{s'})$$

(*ii*) *Let* $\Lambda \subseteq \Psi$ *be a class of formulas,* $A \in \Lambda$ *a formula and* $\vec{t}(\vec{x})$ *are terms such that* $\vec{z} \leq \vec{t}(\vec{x}) \in \Lambda$ *for all variables* $\vec{z}$*. Then if*

$$\mathfrak{B}(\Pi_{k+1},\mathcal{B}) \vdash \forall \vec{x} \exists \vec{y} \leq \vec{t}(\vec{x}) A(\vec{x},\vec{y})$$

then there exists a $\mathrm{BGLS}(\Pi_k,\Lambda,\mathcal{B},\leq)$*-problem with the condition that*

$$G(\vec{x},\vec{y}) = A(\vec{x},\vec{y}) \wedge \vec{y} \leq \vec{t}(\vec{x})$$

Proof. For (i), argue inside $\mathcal{B}$ and assume that there is no $\vec{s'}$ such that $G(\vec{x}, \vec{s'})$. It implies that $\forall \vec{s}(F(\vec{x}, \vec{s}) \to \vec{N}(\vec{x}, \vec{s}) \neq \vec{s})$. Use induction on the formula

$$\forall \vec{s} \leq \vec{r}(\vec{x})[F(\vec{x}, \vec{s}) \to c(\vec{x}, \vec{s}) \geq n]$$

where $\vec{r}(\vec{x})$ is the bound for F. This bound exists because the GLS problem is bounded. First note that the formula is in Π_{k+1}. Hence in $\mathfrak{B}(\Pi_{k+1}, \mathcal{B})$ we can afford such an induction. For $n = 0$ the claim is clear. For $n+1$, assume $F(\vec{x}, \vec{s})$, therefore by the assumption $\vec{N}(\vec{x}, \vec{s}) \neq \vec{s}$ which implies

$$c(\vec{x}, \vec{N}(\vec{x}, \vec{s})) < c(\vec{x}, \vec{s})$$

On the other hand, by $F(\vec{x}, \vec{s})$ we know that $F(\vec{x}, \vec{N}(\vec{x}, \vec{s}))$ and hence $\vec{N}(\vec{x}, \vec{s}) \leq r(\vec{x})$. By IH, we have $c(\vec{x}, \vec{N}(\vec{x}, \vec{s})) \geq n$ which implies $c(\vec{x}, \vec{s}) \geq n + 1$. Therefore, we have

$$\forall n \forall \vec{s} \leq \vec{r}(\vec{x})[F(\vec{x}, \vec{s}) \to c(\vec{x}, \vec{s}) \geq n]$$

Define $c_0 = c(\vec{x}, \vec{i}(\vec{x}))$. For $n = c_0 + 1$ and $\vec{s} = \vec{i}(\vec{x})$ we will have $c_0 \geq c_0 + 1$ which is a contradiction. Hence there exists $\vec{s}$ such that $Goal(\vec{x}, \vec{s})$ which also implies that $\vec{s} \leq \vec{t}(\vec{x})$.

For (ii), assume

$$\mathfrak{B}(\Pi_{k+1}, \mathcal{B}) \vdash \forall \vec{x} \exists \vec{y} \leq \vec{t}(\vec{x}) A(\vec{x}, \vec{y}).$$

Then, we know that $\forall \vec{y} \leq \vec{t}(\vec{x}) \neg A(\vec{x}, \vec{y}) \Rightarrow \bot$ is provable in the theory. Since $A \in \Lambda \subseteq \Pi_k$, we have $\forall \vec{y} \leq \vec{t}(\vec{x}) \neg A(\vec{x}, \vec{y}) \in \Pi_{k+1}$. By soundness theorem 2.4.17, there exist a term $s(\vec{x})$, a formula $H(u, \vec{x}) \in \Pi_{k+1}$ and sequences of terms E_0, E_1, G_0, G_1 and $F(u)$ such that the following statements are provable in $\mathcal{B}$:

(i) $H(0, \vec{x}) \equiv_{\pi}^{(E_0, E_1)} \forall \vec{y} \leq \vec{t}(\vec{x}) \neg A(\vec{x}, \vec{y})$.

(ii) $H(s(\vec{x}), \vec{x}) \equiv_{\pi}^{(G_0, G_1)} \bot$.

(iii) $\forall u < s(\vec{x})\ H(u, \vec{x}) \geq_{\pi}^{F_u} H(u+1, \vec{x})$.

Since $H \in \Pi_{k+1}$, w.l.o.g we can assume $H(u, \vec{x}) = \forall \vec{v} \leq \vec{r}(\vec{x}, u) G(u, \vec{v}, \vec{x})$ where $G(u, \vec{v}, \vec{x}) \in \Sigma_k$ and $\vec{r}$ are monotone. Use the deterministic reductions to show the existence of terms U, V and Z such that

(i) $\mathcal{B} \vdash [\vec{Z}(\vec{x}, \vec{v}) \leq \vec{t}(\vec{x}) \to \neg A(\vec{Z}(\vec{x}, \vec{v}), \vec{x})] \to [\vec{v} \leq \vec{r}(\vec{x}, 0) \to G(0, \vec{v}, \vec{x})]$.

(ii) $\mathcal{B} \vdash [\vec{U}(\vec{x}) \leq \vec{r}(\vec{x}, s(\vec{x})) \to G(s(\vec{x}), \vec{U}(\vec{x}), \vec{x})] \to \bot$.

(iii) $\mathcal{B} \vdash \forall u < s(\vec{x})[\vec{V}(u, \vec{v}, \vec{x}) \leq \vec{r}(\vec{x}, u) \to G(u, \vec{V}(u, \vec{v}, \vec{x}), \vec{x})] \to [\vec{v} \leq \vec{r}(\vec{x}, u + 1) \to G(u+1, \vec{v}, \vec{x})]$.

Now define $B(u, \vec{v}, \vec{z}) = [u \leq s(\vec{x}) \wedge \vec{v} \leq \vec{r}(\vec{x}, s(\vec{x})) \wedge \vec{z} \leq \vec{t}(\vec{x})]$

$$F(\vec{x}; u, \vec{v}, \vec{z}) = \begin{cases} \vec{v} \leq \vec{r}(\vec{x}, u \dot{-} 1) \wedge \neg G(u \dot{-} 1, \vec{v}, \vec{x}) \wedge B(u, \vec{v}, \vec{z}) & u > 0 \\ \vec{z} \leq \vec{t}(\vec{x}) \wedge A(\vec{x}, \vec{z}) \wedge B(u, \vec{v}, \vec{z}) & u = 0 \end{cases}$$

and

$$\vec{N}(\vec{x};u,\vec{v},\vec{z}) = \begin{cases} (u \dot{-} 1, \vec{V}(u,\vec{v},\vec{x}), \vec{z}) & u > 1 \\ (0, \vec{v}, \vec{Z}(\vec{x},\vec{v})) & u = 1 \\ (u,\vec{v},\vec{z}) & u = 0 \end{cases}$$

and $Goal(\vec{x};\vec{z}) = [\vec{z} \leq \vec{t}(\vec{x}) \wedge A(\vec{x},\vec{z})]$, $\vec{i}(\vec{x}) = (s(\vec{x})+1, \vec{U}(\vec{x}), 0)$, and $c(\vec{x};u,\vec{v},\vec{z}) = u$. It is clear to see that this data is a $\mathrm{BGLS}(\Pi_k, \Lambda, \mathcal{B}, \leq)$-problem. The reason is that $F \in \Pi_k$ and $Goal \in \Lambda$ by the assumption. The answer to this problem is $\vec{z}$ such that $\vec{z} \leq \vec{t}$ and $A(\vec{x},\vec{z})$ which completes the proof. □

Corollary 2.4.34.

$$\forall\Sigma_{j+1}[\mathfrak{B}(\Pi_{k+1},\mathcal{B})] \equiv_{\mathcal{B}} \mathrm{BGLS}(\Pi_k,\Pi_j,\mathcal{B},\leq).$$

for all $j \leq k$.

And again the special case for Buss' hierarchy will be:

Corollary 2.4.35. *([6]) For all $j \leq k$, $\forall\Sigma_{j+1}(T_2^{k+1}) \equiv_{\mathrm{PV}} \mathrm{PLS}(\Pi_k,\Pi_j,\mathrm{PV},\leq)$.*

Remark 2.4.36. *Local search problems and the game induction principles provide weaker characterizations than what the theory of flows has to offer. The game induction principle relaxes the $\mathcal{B}$-provability condition of the reductions to make the statement purely combinatorial at the expense of missing some useful information about the provability. The* GLS *problems, though, keep the base theories present, but instead they reduce their reductions to unwind only the outmost block of bounded universal quantifiers, sweeping the rest under the carpet of the feasibility predicate. This is helpful to simplify the formalization, but it clearly misses the huge reduction information that lies in the witnessing of the other quantifiers.*

Using this characterization by the GLS problems, we can also capture the class of all low complexity search problems in strong theories. For the remaining part of this subsection, assume that the languages $\mathcal{L}_{I\Delta_0(\exp)}$ and $\mathcal{L}_{\mathrm{PRA}(\prec)}$ has a separate copy of the language of PV and define $\tilde{\Sigma}_j^b$ and $\tilde{\Pi}_j^b$ as Σ_j and Π_j in the language of PV. For instance, a formula in $\tilde{\Sigma}_1^b$ is essentially in the form $\exists\vec{y} \leq \vec{t}(\vec{x})A(\vec{x},\vec{y})$ where $\vec{t}$ are polynomial-time computable functions and A is a polynomial-time computable predicate. Hence, $\tilde{\Sigma}_1^b$ represents the NP predicates in our greater languages. Moreover, assume that our theories have access to all definitional axioms of PV for their separate language. To emphasize on this modification, we will denote the new version of any theory by the superscript p.

Corollary 2.4.37. *(i)* $\forall\tilde{\Sigma}_{j+1}^b[I\Delta_0^p(\exp)] \equiv_{\mathcal{R}^p(\exp)} \mathrm{BGLS}(\Pi_j,\tilde{\Pi}_j^b,\mathcal{R}^p(\exp),\leq)$.

(ii) $\forall\tilde{\Sigma}_{j+1}^b(\mathrm{PRA}^p) \equiv \mathrm{PLS}((\mathrm{PRA}^-)^p,\leq) \equiv_{(\mathrm{PRA}^-)^p,} \mathrm{BGLS}(\Pi_j,\tilde{\Pi}_j^b,(\mathrm{PRA}^-)^p,\leq)$.

Since we have $\forall\tilde{\Sigma}_{j+1}^b(\mathrm{PRA}_{\prec}^p) \equiv_{\mathrm{BASIC}_{\prec}^p} \mathrm{BGLS}(\Pi_j,\tilde{\Pi}_j^b,\mathrm{BASIC}_{\prec}^p,\leq)$, by the fact that $T \equiv_{\Pi_2^0} \mathrm{PRA}_{\prec_{\alpha_T}}$ we will have:

Theorem 2.4.38. *Let T be a theory of arithmetic with Π_2^0-ordinal α_T with a primitive recursive representation $\prec_{\alpha_T}$, then*

$$\forall\tilde{\Sigma}_{j+1}^b(T^p) \equiv_{\mathrm{BASIC}_{\prec_{\alpha_T}}^p} \mathrm{BGLS}(\Pi_j,\tilde{\Pi}_j^b,\mathrm{BASIC}_{\prec_{\alpha_T}}^p,\leq)$$

Corollary 2.4.39. *Let $\epsilon(\alpha)$ be the least ϵ number after α with a primitive recursive representation. Then*

$$\forall\tilde{\Sigma}^b_{j+1}([\mathrm{PA}+\mathrm{TI}(\alpha)]^p) \equiv_{\mathrm{BASIC}^p_{\prec_{\epsilon(\alpha)}}} \mathrm{BGLS}(\Pi_j, \tilde{\Pi}^b_j, \mathrm{BASIC}^p_{\prec_{\epsilon(\alpha)}}, \leq)$$

Remark 2.4.40. *These characterizations of the low complexity consequences of the strong theories of arithmetic may seem a bit counter-intuitive. The reason is the paradoxical situation in which we have full access to a class of extremely complex functions while the search problems that we try to solve are much easier. A typical example of such a mismatch is our characterization of the total $\tilde{\Sigma}^b_1 = \mathrm{NP}$ search problems of the theory $I\Delta_0(\exp)$. What the Lemma 2.4.37 presents is an algorithm based on a sequence of elementary computable reductions, while our* NP *search problem is just a very low complexity problem solvable by a brute force search in exponential time. Based on this mismatch, it may seem natural to conclude the sufficiency of one obvious reduction which implies the triviality of our characterizations. This is not a sound argument. It is correct that we have full access to a certain class of complex functions but it does not mean that we have full access to their complete theory about their behavior. What we know is usually a very basic theory consisting of the defining axioms of the function symbols. These complex functions behave as oracles to which we can impose our questions, but we can't fully understand their behavior, and hence we can't be sure about the correctness of their computations. Here is where the long sequences of reductions come to rescue. They consist of very simple computational steps based on the definitional axioms of the functions so that in each reduction we can ensure that our computation works correctly. In fact, reductions decompose a computation to simple verifiable steps which actually simulates the application of the induction axiom in the proof of the totality of the search problem.*

2.5 Ordinal Flows

In the previous sections we have developed a theory for deterministic and non-deterministic flows to investigate provability in the bounded theories of arithmetic. In this section we will generalize the theory to embrace also some stronger unbounded theories. For this purpose, we will pursue the following path: Since we are interested just in the low-complexity consequences of the theories, we will first use the continuous cut elimination technique to transfer these consequences of the theories to a simpler theory axiomatized by transfinite induction on the universal statements. Then we will extend the length of the flows from terms to infinite ordinals to be able to deal with these long ordinal-length inductions.

To implement this task we need the following polynomial time representation of the very basic ordinal arithmetic. Note that our goal is importing the ordinal analysis of the given theories (usually done over PRA) from the primitive recursive setting to the polynomial time setting. This helps skipping the whole process of reimplementation of the ordinal analysis over PV. For this purpose, from now on, we will assume that the ordinal and its basic arithmetic are given by their fixed primitive recursive representation.

Definition 2.5.1. *Let α be an ordinal with a primitive recursive representation. Then we say*

$$\mathbb{A} = (A, \prec_A, +_A, \cdot_A, \dot{-}_A, d_A(\cdot,\cdot), 0_A, 1_A)$$

is a polytime representation of the ordinal α when A and $\prec_A$ are polytime relations, $+_A, \cdot_A, \dot{-}_A, d_A(\cdot,\cdot)$ are polytime functions and constants $0_A, 1_A$ such that:

(*i*) *The structure $\mathbb{A} = (A, \prec_A, +_A, \cdot_A, \dot{-}_A, d_A(\cdot,\cdot), 0_A, 1_A)$ is isomorphic to $\mathfrak{A} = (\alpha, \prec_\alpha, +_\alpha, \cdot_\alpha, \dot{-}_\alpha, d_A(\cdot,\cdot), 0_\alpha, 1_\alpha)$ where $\dot{-}_\alpha, d_A(\cdot,\cdot)$ are subtraction and division from left, i.e. for $\beta \preceq \alpha$ we have $\alpha \dot{-} \beta = \gamma$ where $\beta + \gamma = \alpha$ and otherwise, $\alpha \dot{-} \beta = 0$. For division, if $\beta \neq 0$, by $d(\alpha, \beta)$ we mean the unique γ where $\alpha = \beta\gamma + \delta$ and $\delta \prec \beta$.*

(*ii*) PV *proves the axioms of discrete ordered semi-rings for the structure $\mathbb{A}$ without the commutativity of addition and the axioms which state that $\prec_A$ preserves under left addition and left multiplication by a non-zero element.*

(*ii*) PRA *proves that $\mathbb{A}$ is equivalent to the primitive recursive representation of $\mathfrak{A}$.*

Definition 2.5.2. *Let $\mathcal{L}_{\mathrm{PV}}$ be the language of* PV. *Define the system* $\mathrm{TI}(\forall_1, \prec)$ *as the usual first order sequent calculus of first order language plus the axioms of* PV *and the following induction rule for any a:*

$$(Ind_\alpha)\ \frac{\Gamma, \delta \prec a, \forall\gamma \prec_a \delta\ A(\gamma) \Rightarrow \Delta, A(\delta)}{\Gamma, \theta \prec a \Rightarrow \Delta, A(\theta)}$$

such that every formula in the proof just consists of formulas in the class $\forall_1$ where $\forall_1$ means the class of all universal formulas which is inductively defined as the least set that includes atomic formulas and is closed under conjunction, disjunction, implication with quantifier-free precedent and universal quantifiers, a is the code for the ordinal α and $\prec_a$ means the order $\prec$ on the set $\{b | b \prec a\}$.

Remark 2.5.3. *For some practical reasons, it is useful to change the induction rule to the rule:*

$$(Ind'_\alpha)\ \frac{\Gamma, \delta \prec a, \forall\gamma \prec_a \delta\ A(\gamma) \Rightarrow \Delta, \forall\gamma \prec_a \delta + 1\ A(\gamma)}{\Gamma, \theta \prec a, \Rightarrow \Delta, A(\theta)}$$

In the presence of the other first order rules specifically the $\forall_1$-cut rule, the equivalence of these two induction rules is trivial.

Recall the following definition of the Π^0_2-proof theoretical ordinal of a theory T:

Definition 2.5.4. *Let T be a theory of arithmetic. We say that α is a Π^0_2-proof theoretical ordinal of T when $\prec$ is the primitive recursive representation of α and $T \equiv_{\Pi^0_2} \mathrm{PRA} + \bigcup_{\beta \prec \alpha} \mathrm{TI}(\prec_\beta)$, where $\mathrm{TI}(\prec_\beta)$ means full transfinite induction up to the ordinal β.*

Convention. From now on wherever we have a proof theoretic ordinal, we always assume that it is closed under the operation $\beta \mapsto \omega^\beta$.

The following theorem uses continuous cut elimination technique to reduce full transfinite induction to simpler one.

Theorem 2.5.5. *[11] Let T be a theory of arithmetic and α its Π^0_2-proof theoretical ordinal. Then*

$$T \equiv_{\Pi^0_2} \mathrm{PRA} + \bigcup_{\beta \prec \alpha} \forall_1 - \mathrm{TI}(\prec_\beta)$$

Using the Definition 2.5.4 and Theorem 2.5.5, we can transfer Π^0_2 consequences of the theory T to the theory $\mathrm{PRA} + \bigcup_{a \in A} \forall_1 - \mathrm{TI}(\prec_a)$ where $\prec_a$ is a primitive recursive representation of α_T up to a. The following lemma makes it possible to continue this process of transferring to reach the theory $\mathrm{TI}(\forall_1, \prec)$ which is more convenient for our technical purposes.

Lemma 2.5.6. $\mathrm{PRA} + \bigcup_{a \in A} \forall_1 - \mathrm{TI}(\prec_a)$ *is a $\forall_1$-sequent conservative extension of the theory $\mathrm{TI}(\forall_1, \prec)$, i.e., for any $\Gamma(\vec{x}) \cup \Delta(\vec{x}) \subseteq \forall_1$, $\mathrm{TI}(\forall_1, \prec) \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$ iff*

$$\mathrm{PRA} + \bigcup_{a \in A} \forall_1 - \mathrm{TI}(\prec_a) \vdash \Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$$

Proof. One direction is trivial. For the other direction, first note that the usual translation of all quantifier-free formulas of the language of PRA into the $\Delta_1(I\Pi^0_1)$ statements in the language of Peano arithmetic to map PRA into $I\Pi^0_1$, can be slightly changed to map the quantifier-free formulas into the class $\forall_1$. (Use the unbounded quantifiers as the witness for the whole computation of the involving primitive recursive functions.) Extending this process to our languages both of which includes the language of PV, we can assign a $\forall_1$ formula to any quantifier-free formula in the language of PRA and since the primitive recursive representation of the ordinal is equivalent to its polytime representation provably in PRA, we can interpret the theory $\mathrm{PRA}+\bigcup_{a \in A} \forall_1 - \mathrm{TI}(\prec_a)$ into the theory $\mathrm{TI}(\forall_1, \prec)$. □

Corollary 2.5.7. *Let α be the Π^0_2-ordinal of the theory T, then we have $T \equiv_{\forall_1} \mathrm{TI}(\forall_1, \prec)$ in the sequent sense of the Lemma 2.5.6.*

We have defined our theory so far. Let us now define the concept of an ordinal flow:

Definition 2.5.8. *Let $A(\vec{x})$, $B(\vec{x})$ and $H(\delta, \vec{x})$ be some formulas in $\forall_1$. A tuple (H, β) where $\beta \prec \alpha$ is called an α-flow if*

(i) $\mathrm{PV} \vdash A(\vec{x}) \leftrightarrow H(0, \vec{x})$.

(ii) $\mathrm{PV} \vdash \forall\, 1 \preceq \delta \prec \beta\ [\forall \gamma \prec \delta\ H(\gamma, \vec{x}) \rightarrow \forall \gamma \prec \delta + 1\ H(\gamma, \vec{x})]$.

(iii) $\mathrm{PV} \vdash H(\beta, \vec{x}) \leftrightarrow B(\vec{x})$.

We denote the existence of an α-flow from A to B by $A \rhd_\alpha B$ and we abbreviate $\bigwedge \Gamma \rhd_\alpha \bigvee \Delta$ by $\Gamma \rhd_\alpha \Delta$. Moreover, when it is clear from the context, we omit the subscript α everywhere.

Like in the bounded case, we need to prove some basic lemmas for this new notion of ordinal flow. These lemmas then help to prove the corresponding soundness theorem.

Lemma 2.5.9. *(Conjunction Application) Let $C(\vec{x}) \in \forall_1$ be a formula. If $A(\vec{x}) \rhd B(\vec{x})$ then $A(\vec{x}) \wedge C(\vec{x}) \rhd B(\vec{x}) \wedge C(\vec{x})$.*

Proof. Since $A(\vec{x}) \rhd B(\vec{x})$, then by Definition 2.5.8 there exist an ordinal β and a formula $H(\gamma, \vec{x}) \in \forall_1$ such that we have the conditions in the Definition 2.5.8. Define $\beta' = \beta$ and $H'(\gamma, \vec{x}) = H(\gamma, \vec{x}) \wedge C(\vec{x})$. It is clear that the (H', β') is an α-flow from $A(\vec{x}) \wedge C(\vec{x})$ to $B(\vec{x}) \wedge C(\vec{x})$. □

Lemma 2.5.10. *(Disjunction Application) Let $C(\vec{x}) \in \forall_1$ be a formula. If $A(\vec{x}) \rhd B(\vec{x})$ then $A(\vec{x}) \vee C(\vec{x}) \rhd B(\vec{x}) \vee C(\vec{x})$.*

Proof. Since $A(\vec{x}) \rhd B(\vec{x})$, then by Definition 2.5.8, there exist an ordinal β and a formula $H(\gamma, \vec{x}) \in \forall_1$ such that the conditions in the Definition 2.5.8 is provable in PV. Now define $\beta' = \beta$ and $H'(\gamma, \vec{x}) = H(\gamma, \vec{x}) \vee C(\vec{x})$. It is easy to see that (H', β') is an α-flow from $A(\vec{x}) \vee C(\vec{x})$ to $B(\vec{x}) \vee C(\vec{x})$. □

Lemma 2.5.11. *(i) (Weak Gluing) If $A(\vec{x}) \rhd B(\vec{x})$ and $B(\vec{x}) \rhd C(\vec{x})$, then $A(\vec{x}) \rhd C(\vec{x})$.*

(ii) (Strong Gluing) If $\forall \gamma \prec \delta \; A(\gamma, \vec{x}) \rhd \forall \gamma \prec \delta + 1 \; A(\gamma, \vec{x})$, then $\rhd A(\theta, \vec{x})$.

Proof. For (i), since $A(\vec{x}) \rhd B(\vec{x})$ there exist an ordinal β and a formula $H(\gamma, \vec{x}) \in \forall_1$ such that PV proves the conditions in the Definition 2.5.8. On the other hand since $B(\vec{x}) \rhd C(\vec{x})$ we have the corresponding data for $B(\vec{x})$ to $C(\vec{x})$ which we show by β' and $H'(\gamma, \vec{x})$. Define $\beta'' = \beta + \beta'$ and

$$H''(\gamma, \vec{x}) = \begin{cases} H(\gamma, \vec{x}) & \gamma \preceq \beta \\ H'(\gamma \dot{-} \beta, \vec{x}) & \beta \prec \gamma \preceq \beta + \beta' \end{cases}$$

It is easy to check that (β'', H'') is an α-flow from $A(\vec{x})$ to $C(\vec{x})$.

For (ii) first let us prove $\forall \gamma \prec 0 \; A(\gamma, \vec{x}) \rhd \forall \gamma \prec \theta + 1 \; A(\gamma, \vec{x})$. If we have $\forall \gamma \prec \delta A(\gamma, \vec{x}) \rhd \forall \gamma \prec \delta + 1 \; A(\gamma, \vec{x})$ then there exists β and $H(\eta, \delta, \vec{x})$ such that we have the conditions of the Definition 2.5.8. Define $\beta' = \beta \times (\theta + 1)$ and $I(\tau, \vec{x}) = H(\tau \dot{-} \beta d(\tau, \beta), d(\tau, \beta), \vec{x})$. It is easy to see that (I, β') is an α-flow from $\forall \gamma \prec 0 \; A(\gamma, \vec{x})$ to $\forall \gamma \prec \theta + 1 \; A(\gamma, \vec{x})$. Now it is enough to add $A(\theta)$ in the end of the flow and it completes the proof. □

Lemma 2.5.12. *(Conjunction and Disjunction Rules)*

(i) If $\Gamma, A \rhd \Delta$ or $\Gamma, B \rhd \Delta$, then $\Gamma, A \wedge B \rhd \Delta$.

(ii) If $\Gamma_0 \rhd \Delta_0, A$ and $\Gamma_1 \rhd \Delta_1, B$, then $\Gamma_0, \Gamma_1 \rhd \Delta_0, \Delta_1, A \wedge B$.

(iii) If $\Gamma \rhd \Delta, A$ or $\Gamma \rhd \Delta, B$, then $\Gamma \rhd \Delta, A \vee B$.

(iv) If $\Gamma_0, A \rhd \Delta_0$ and $\Gamma_1, B \rhd \Delta_1$, then $\Gamma_0, \Gamma_1, A \vee B \rhd \Delta_0, \Delta_1$.

Proof. The proof is similar to the proof of the Lemma 2.3.14. Note that the proof of the Lemma 2.3.14 is fully based on the weak gluing and applying conjunction and disjunction everywhere in the flows, which means that we can apply the same proof wherever we have those properties. □

Theorem 2.5.13. *(Soundness) If $\Gamma \cup \Delta \subseteq \forall_1$ and $\mathrm{TI}(\forall_1, \prec) \vdash \Gamma \Rightarrow \Delta$, then there exists an α-flow from Γ to Δ.*

Proof. We prove the lemma by induction on the length of the proof of $\Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$ using the induction rule mentioned in the Remark 2.5.3. Note that the proof consists only of $\forall_1$ formulas by definition.

1. (Axioms). If $\Gamma(\vec{x}) \Rightarrow \Delta(\vec{x})$ is a logical axiom then the claim is trivial. If it is a non-logical axiom then the claim will be also trivial because all non-logical axioms are provable in PV. Therefore there is nothing to prove.

2. (Structural Rules). These are derivable from the same rules available in PV.

3. (Cut). See the Remark 2.3.16.

4. (Propositional Rules). The conjunction and disjunction cases are proved in the Lemma 2.5.12. The implication and negation cases are easy because they should have quantifier-free precedents and be quantifier-free, respectively, and hence we can manipulate them as in the Lemma 2.3.15 and 2.3.17.

5. (Universal Quantifier Rules, Right). If $\Gamma(\vec{x}) \Rightarrow \Delta(\vec{x}), \forall z B(\vec{x}, z)$ is proved by the $\forall R$ rule by $\Gamma(\vec{x}) \Rightarrow \Delta(\vec{x}), B(\vec{x}, b)$, then by IH, $\Gamma(\vec{x}) \rhd \Delta(\vec{x}), B(\vec{x}, b)$. Therefore, there exist an ordinal β and a formula $H(\gamma, \vec{x}, b) \in \forall_1$ such that the conditions of the Definition 2.5.8 are provable in PV. Define $\beta' = \beta$ and $H'(\gamma, \vec{x}) = \forall b H(\gamma, \vec{x}, b)$. Since $H(\gamma, \vec{x}, b) \in \forall_1$ we have $\forall b H(\gamma, \vec{x}, b) \in \forall_1$. The other conditions to ensure that the new sequence is an α-flow from $\forall b[\bigwedge \Gamma(\vec{x})]$ to $\forall b[B(\vec{x}, b) \vee \bigvee \Delta]$ is a straightforward consequence of the fact that if

$$\mathrm{PV} \vdash \forall \gamma \prec \delta H(\gamma, b, \vec{x}) \to \forall \gamma \prec \delta + 1 \; H(\gamma, b, \vec{x}),$$

then

$$\mathrm{PV} \vdash \forall \gamma \prec \delta \forall b H(\gamma, z, \vec{x}) \to \forall \gamma \prec \delta + 1 \forall b H(\gamma, z, \vec{x}).$$

Finally, note that $\Gamma \cup \Delta$ does not have a free b variable and hence $\forall b[\bigwedge \Gamma]$ and $\forall b[B(\vec{x}, z) \vee \bigvee \Delta]$ are equivalent to $\bigwedge \Gamma$ and $\bigvee \Delta \vee \forall b B(\vec{x}, z)$, provably in PV and since changing the name of a bounded variable from b to z does not change the nature of a flow, we can complete the proof.

6. (Universal Quantifier Rules, Left). If $\Gamma(\vec{x}), \forall z B(\vec{x}, z) \Rightarrow \Delta(\vec{x})$ is proved by the $\forall L$ rule by $\Gamma(\vec{x}), B(\vec{x}, s(\vec{x})) \Rightarrow \Delta(\vec{x})$, then since $\mathrm{PV} \vdash \forall z B(\vec{x}, z) \to B(\vec{x}, s(\vec{x}))$, and

$$\Gamma(\vec{x}), B(\vec{x}, s(\vec{x})) \rhd \Delta(\vec{x}),$$

we have

$$\Gamma(\vec{x}), \forall z B(\vec{x}, z) \rhd \Delta(\vec{x}).$$

7. (Induction). The proof is a combination of the strong gluing 2.5.11 and some propositional maneuver mentioned in the Remark 2.3.16.

□

Theorem 2.5.14. *(Completeness) If* $\Gamma \cup \Delta \subseteq \forall_1$ *and* $\Gamma \rhd \Delta$*, then* $\mathrm{TI}(\forall_1, \prec) \vdash \Gamma \Rightarrow \Delta$.

Proof. If there exists an α-flow from Γ to Δ then it means that there exists (H, β) such that

(*i*) $\mathrm{PV} \vdash \bigwedge \Gamma(\vec{x}) \leftrightarrow H(0, \vec{x})$.

(*ii*) $\mathrm{PV} \vdash \forall\, 1 \preceq \delta \prec \beta\ [\forall \gamma \prec \delta\ H(\gamma, \vec{x}) \rightarrow \forall \gamma \prec \delta + 1\ H(\gamma, \vec{x})]$.

(*iii*) $\mathrm{PV} \vdash H(\beta, \vec{x}) \leftrightarrow \bigvee \Delta(\vec{x})$.

Therefore, using induction on $H(\delta, \vec{x})$ we have

$$\mathrm{TI}(\forall_1, \prec) \vdash H(0, \vec{x}) \Rightarrow H(\beta, \vec{x}),$$

and thus $\mathrm{TI}(\forall_1, \prec) \vdash \bigwedge \Gamma(\vec{x}) \Rightarrow \bigvee \Delta(\vec{x})$. □

Corollary 2.5.15. *Let T be a theory of arithmetic, α_T be its proof theoretic ordinal with a polynomial time representation and $\Gamma \cup \Delta \subseteq \forall_1$. Then $T \vdash \Gamma \Rightarrow \Delta$ iff $\Gamma \rhd_{\alpha_T} \Delta$.*

In the following we will use ordinal PLS problems to characterize the NP search problems of any theory with proof theoretic ordinal α.

Theorem 2.5.16. *Let T be a theory of arithmetic and α_T be its proof theoretic ordinal with a polynomial time representation, then* $\mathrm{TFNP}(T) \equiv_{\mathrm{PV}} \bigcup_{\beta \prec \alpha_T} \mathrm{PLS}(\prec_\beta)$.

Proof. First, note that all PLS problems are provably total in

$$\mathrm{PRA} + \bigcup_{a \in A} \mathrm{TI}(\forall_1, \prec_a)$$

To prove this fact, first find the least ordinal β such that $\exists \vec{s}\ (c(\vec{x}, \vec{s}) = \beta \wedge F(\vec{x}, \vec{s}))$. Its existence is a consequence of an instance of a $\forall_1$-transfinite induction and the existence of at least one ordinal, namely $c(\vec{x}, \vec{i}(\vec{x}))$, for which $\exists \vec{s}\ (c(\vec{x}, \vec{s}) = c(\vec{x}, \vec{i}(\vec{x})) \wedge F(\vec{x}, \vec{s}))$ holds. The corresponding $\vec{s}$ for this minimal β is the one whose projection is the answer for the PLS problem because if not, $N(\vec{x}, \vec{s})$ should have a lower cost which contradicts the choice of β. Hence, $\mathrm{PRA} + \bigcup_{a \in A} \mathrm{TI}(\forall_1, \prec_a)$ proves the totality of the PLS problem and since it is a Π^0_1 statement, its provability transfers also to the theory T.

For the converse, assume that $T \vdash \forall \vec{x} \exists \vec{y}\ [|\vec{y}| \leq \vec{p}(|\vec{x}|) \wedge A(\vec{x}, \vec{y})]$ where $A(\vec{x}, \vec{y})$ is quantifier-free in the language of PV. Then by the Corollary 2.5.15 we have $\forall \vec{y}(|\vec{y}| \leq \vec{p}(|\vec{x}|) \rightarrow \neg A(\vec{x}, \vec{y})) \rhd \bot$. Hence there exists (H, β) such that

(*i*) $\mathrm{PV} \vdash \forall \vec{y}(|\vec{y}| \leq \vec{p}(|\vec{x}|) \rightarrow \neg A(\vec{x}, \vec{y})) \rightarrow H(0, \vec{x})$.

(*ii*) $\mathrm{PV} \vdash \forall\, 1 \preceq \gamma \prec \beta\ [\forall \delta \prec \gamma\ H(\delta, \vec{x}) \rightarrow \forall \delta \prec \gamma + 1\ H(\delta, \vec{x})]$.

(*iii*) $\mathrm{PV} \vdash H(\beta, \vec{x}) \rightarrow \bot$.

Since $H \in \forall_1$ we have $H(\gamma, \vec{x}) \equiv_{\mathrm{PV}} \forall \vec{z} G(\gamma, \vec{x}, \vec{z})$ where G is quantifier-free. On the other hand, all the conditions are provable in PV which means that we can witness the existential quantifiers by polytime functions. Hence, there are polytime functions $\vec{Y}(\vec{x}, \vec{z})$, $\vec{Z}(\vec{x}, \vec{z}, \delta)$, $\Delta(\vec{x}, \vec{z}, \delta)$ and $\vec{W}(\vec{x})$ such that:

(i') $\mathrm{PV} \vdash (|\vec{Y}(\vec{x},\vec{z})| \leq p(|\vec{x}|) \to \neg A(\vec{x},\vec{Y}(\vec{x},\vec{z}))) \to G(0,\vec{x},\vec{z})$.

(ii') $\mathrm{PV} \vdash \forall\ 1 \preceq \gamma \prec \beta\ [\Delta(\vec{x},\vec{z},\delta) \prec \gamma \to G(\Delta(\vec{x},\vec{z},\delta),\vec{x},\vec{Z}(\vec{x},\vec{z},\delta)) \to \delta \prec \gamma + 1 \to G(\delta,\vec{x},\vec{z})]$.

(iii') $\mathrm{PV} \vdash G(\beta,\vec{x},\vec{W}(\vec{x})) \to \bot$.

Put $\delta = \gamma$ in (ii'), then we have

$$\mathrm{PV} \vdash \forall \gamma \prec \beta\ [(\Delta(\vec{x},\vec{z},\gamma) \prec \gamma \to G(\Delta(\vec{x},\vec{z},\gamma),\vec{x},\vec{Z}(\vec{x},\vec{z},\gamma)) \to G(\gamma,\vec{x},\vec{z})].$$

Define

$$F(\vec{x};\gamma,\vec{y},\vec{z}) = \begin{cases} \neg G(\vec{x},\gamma,\vec{z}) & \text{if } \omega \preceq \gamma \\ \neg G(\vec{x},\gamma \dot{-} 1,\vec{z}) & \text{if } 0 \prec \gamma \prec \omega \\ |\vec{y}| \leq \vec{p}(|\vec{x}|) \wedge A(\vec{x},\vec{y}) & \text{if } \gamma = 0 \end{cases}$$

and

$$N(\vec{x};\gamma,\vec{y},\vec{z}) = \begin{cases} (\Delta(\vec{x},\vec{z},\gamma),\vec{y},\vec{Z}(\vec{x},\vec{z},\gamma)) & \omega \preceq \gamma, \neg G(\vec{x},\gamma,\vec{z}) \\ (0,\vec{y},\vec{0}) & \omega \preceq \gamma, G(\vec{x},\gamma,\vec{z}) \\ (\Delta(\vec{x},\vec{z},\gamma \dot{-} 1) + 1,\vec{y},\vec{Z}(\gamma \dot{-} 1)) & 1 \prec \gamma \prec \omega, \neg G(\vec{x},\gamma \dot{-} 1,\vec{z}) \\ (0,\vec{y},\vec{0}) & 1 \prec \gamma \prec \omega, G(\vec{x},\gamma \dot{-} 1,\vec{z}) \\ (0,\vec{Y}(\vec{x},\vec{z}),\vec{z}) & \gamma = 1 \\ (\gamma,\vec{y},\vec{z}) & \gamma = 0 \end{cases}$$

and $i(\vec{x}) = (\vec{x},\beta,\vec{0},\vec{W}(\vec{x}))$ and $c(\vec{x};\gamma,\vec{y},\vec{z}) = \gamma$,

$$\mathit{Goal}(\vec{x};\vec{y}) = |\vec{y}| \leq \vec{p}(|\vec{x}|) \wedge A(\vec{x},\vec{y})$$

It is easy to see that this new data is a $\mathrm{PLS}(\prec_{\beta+1})$ problem and its answer is $\vec{y}$ where $|\vec{y}| \leq \vec{p}(|\vec{x}|) \wedge A(\vec{x},\vec{y})$. □

We believe that the notation system introduced in [5] actually provides a polytime representation of the ordinal ϵ_0. Given this fact, as a corollary we will have:

Corollary 2.5.17. *(i) ([3])* $\mathrm{TFNP(PA)} \equiv_{\mathrm{PV}} \bigcup_{\beta \prec \epsilon_0} \mathrm{PLS}(\prec_\beta)$.

(ii) Let α be an ordinal and $\epsilon(\alpha)$ be the least ϵ number after α with a polynomial-time representation. Then

$$\mathrm{TFNP(PA + TI}(\alpha)) \equiv_{\mathrm{PV}} \bigcup_{\beta \prec \epsilon(\alpha)} \mathrm{PLS}(\prec_\beta)$$

Bibliography

[1] M. Ardeshir, B. Hesaam, An introduction to Basic Arithmetic, Logic Jnl IGPL (2008) 16 (1): 1-13.

[2] S. Artemov, Explicit provability and constructive semantics. Bulletin of Symbolic Logic, 7(1):1-36, 2001.

[3] A. Beckmann, A Characterisation of Definable NP Search Problems in Peano Arithmetic, Logic, Language, Information and Computation, 16th International Workshop, WoLLIC 2009, Tokyo, Japan, June 21-24, 2009.

[4] A. Beckmann and S. R. Buss, Characterization of Definable Search Problems in Bounded Arithmetic via Proof Notations, Ontos Verlag, 2010, pp. 65-134.

[5] A. Beckmann, S. R. Buss, C. Pollett, Ordinal Notations and Well-Orderings in Bounded Arithmetic, Annals of Pure and Applied Logic 120(2002), 197-223.

[6] A. Beckmann and S. R. Buss, Polynomial local search in the polynomial hierarchy and witnessing in fragments of bounded arithmetic, Journal of Mathematical Logic, 9 (2009), pp. 103-138.

[7] G. Boolos, The logic of provability, Cambridge University Press, 1993.

[8] S. R. Buss, Bounded Arithmetic, Bibliopolis, Naples, Italy, 1986.

[9] S. R. Buss, The modal logic of pure provability, Notre Dame Journal of Formal Logic, vol. 31 (1990), no. 2, pp. 225-231.

[10] A. Chagrov and M. Zakharyaschev, Modal Logic, Oxford University Press, 1997.

[11] H. Friedman and S. Sheard, Elementary descent recursion and proof theory, Annals of Pure and Applied Logic 71 (1995) 1–45.

[12] K. Gödel, Eine Interpretation des Intuitionistichen Aussagenkalküls, Ergebnisse Math Colloq. Vol. 4 (1933), pp. 39-40.

[13] L. A. Kolodziejczyk, P. Nguyen, and N. Thapen, The provably total NP search problems of weak second-order bounded arithmetic, Annals of Pure and Applied Logic, 162 (2011).

[14] J. Krajicek, P. Pudlak, G. Takeuti, Bounded arithmetic and the polynomial hierarchy, Annals of Pure and Applied Logic, 52: 143-53.

[15] S. Kripke, Semantical considerations on modal logic, Acta Philosophica Fennica, vol. 16 (1963), pp. 83-94.

[16] F. Poggiolesi, Gentzen Calculi for Modal Propositional Logic, Springer, 2010.

[17] W. Ruitenburg, Basic logic and Fregean set theory. In H. Barendregt, M. Bezem, J.W. Klop (editors). Dirk van Dalen Festschrift. Quaestiones In nitae Vol. 5, Department of Philosophy, Utrecht University, 1993, 121-142.

[18] H. Schwichtenberg, A. Troelstra, Basic Proof Theory, Second Edition, Cambridge University Press, 2000.

[19] A. Skelley and N. Thapen, The provably total search problems of bounded arithmetic, Proceedings of the London Mathematical Society, 103 (2011), pp. 106-138.

[20] R. Solovay, Provability interpretations of modal logic, Israel Journal of Mathematics, vol. 25 (1976), pp. 287-304.

[21] N. Thapen, Higher complexity search problems for bounded arithmetic and a formalized no-gap theorem, Archive for Mathematical Logic, Vol 50:7-8, pages 665-680, 2011.

[22] A. Visser, A propositional logic with explicit fixed points. Studia Logica 40 (1981), 155-175.