On the Deduction Rule and the Number of Proof Lines

Maria Luisa Bonet*

Department of Mathematics
U.C. Berkeley
Berkeley, California 94720

Samuel R. Buss*

Department of Mathematics
U.C. San Diego
La Jolla, California 92093

Abstract

We introduce new proof systems for propositional logic, *simple deduction Frege systems*, *general deduction Frege systems* and *nested deduction Frege systems*, which augment Frege systems with variants of the deduction rule. We give upper bounds on the lengths of proofs in these systems compared to lengths in Frege proof systems. As an application we give a near-linear simulation of the propositional Gentzen sequent calculus by Frege proofs. The length of a proof is the number of steps or lines in the proof.

A general deduction Frege proof system provides at most quadratic speedup over Frege proof systems. A nested deduction Frege proof system provides at most a nearly linear speedup over Frege system where by "nearly linear" is meant the ratio of proof lengths is $O(\alpha(n))$ where α is the inverse Ackermann function. A nested deduction Frege system can linearly simulate the propositional sequent calculus and hence a Frege proof system can simulate the propositional sequent calculus with proof lengths bounded by $O(n \cdot \alpha(n))$.

As a technical tool, we introduce the serial transitive closure problem: given a directed graph and a list of *closure edges* in the transitive closure of the graph, the problem is to derive all the closure edges. We give a nearly linear bound on the number of steps in such a derivation when the graph is tree-like.

1 Introduction

A Frege proof system is an inference system for propositional logic in which the only rule of inference is modus ponens. Although it suffices to have modus ponens as the single inference rule to obtain a complete proof system, it is well-known that other modes of inference are also sound. A notable example of this is the deduction rule which states that if a formula B has a proof from an additional, extra-logical hypothesis A (in symbols, $A \vdash B$) then there is a proof of $A \supset B$. This paper considers various strengthenings of this deduction rule and establishes upper bounds on the proof-speedups obtained with these deduction rules.

By a "speedup" of a proof, we mean the amount that proofs can be shortened with additional inference rules. In this paper, the *length* of a proof is the number of lines in the proof; where a line consists either a formula or a sequent (depending on the proof system). We write $\vdash_{\overline{k}} B$ (and $A_1, \ldots, A_s \vdash_{\overline{k}} B$) to indicate that the formula B has Frege proof of $\leq k$ lines (from the hypotheses $A_1, \ldots, A_s$). More generally, we write "$\vdash^{T}_{\overline{k}}$" to mean "provable in proof system T with $\leq k$ lines". If S and T are proof systems we say that S can linearly (respectively, quadraticly) simulate T if, for any T-proof of k lines, there is an S-proof of the same (or sometimes an equivalent) formula of $O(k)$ lines (respectively, of $O(k^2)$ lines). We say that T provides at most linear (respectively, quadratic speedup) over S if S can linearly (respectively, quadraticly) simulate T.

An alternative, commonly used measure of the length of a propositional proof is the number of symbols in the proof. This is the approach used, for instance, by Cook-Reckhow [3] and Statman [12]. It should be noticed that the minimum number of lines in a Frege proof of a formula is polynomially related to the minimum number of symbols in an extended Frege proof. On the other hand, prior work on proof lengths in first-order logic has frequently measured the number of lines in proofs; this includes [2,4,6,7,8,10,11] and others.

We begin by defining the main propositional proof systems used in this paper. The logical connectives of all our systems are presumed to be $\neg$, $\lor$, $\land$ and $\supset$; however, our results hold for any complete set of connectives.

Definition A *Frege proof system* (denoted $\mathcal{F}$) is characterized by:

(1) A finite set of axiom schemata. For example, a

* Supported in part by NSF Grant DMS-8902480

possible axiom schema is $(A \supset (B \supset A))$; A and B represent arbitrary formulas.

(2) The only rule of inference is Modus Ponens (MP):

$$\frac{A \qquad A \supset B}{B}$$

(3) A proof in this system is a sequence of formulas $A_1, \ldots, A_n$ (also called 'lines') where each A_i is either a substitution instance of an axiom schema or is inferred by Modus Ponens (MP) from some A_j and A_k with $j, k < i$.

(4) The proof system must be consistent and complete.

The *length* of an $\mathcal{F}$-proof is the number of lines in the proof; we write $\vdash_{\overline{k}} A$ to indicate that A has a Frege proof of length $\leq k$. We further write $A_1, \ldots, A_s \vdash_{\overline{k}} B$ to mean that B is provable from the hypotheses A_i with a Frege proof of $\leq k$ lines; in other words, that there is a sequence of $\leq k$ formulas each of which is one of the A_i's, is an axiom, or is inferred by modus ponens from earlier formulas such that B is the final formula of the proof. Although we have not specified the axiom schemata to be used in a Frege proof system, it is easy to see that different choices of axiom schemata will change the lengths of proofs only linearly.

The simplest form of the deduction theorem states that if $A \vdash B$ then $\vdash A \supset B$. This can be informally phrased as a rule in the form

$$\frac{A \vdash B}{A \supset B}$$

which is called the *1-simple deduction rule*; more generally, the *simple deduction rule* is

$$\frac{A_1, \ldots A_n \vdash B}{A_1 \supset (\cdots (A_n \supset B) \cdots)}$$

We next define extensions to the Frege proof system that incorporate the deduction theorem as a rule of inference. For this purpose, the systems defined below have proofs in which the lines are *sequents* of the form $\Gamma \models A$; intuitively, the sequent means that the formulas in Γ tautologically imply A: operationally, a sequent $\Gamma \models A$ means that A has been proved using the formulas in Γ as assumptions.

A *general deduction Frege system* (denoted $d\mathcal{F}$) incorporates a strong version of the deduction rule. Each line in a general deduction Frege proof is a sequent of the form $\Gamma \models A$ where A is a formula and Γ is a set of formulas. When Γ is empty, we write just $\models A$. The four valid axioms and inference rules in a general deduction Frege proof are:

$\models A \quad - A$ an axiom

$\{A\} \models A \quad$ – Hypothesis

$$\frac{\Gamma_1 \models A \supset B \qquad \Gamma_2 \models A}{\Gamma_1 \cup \Gamma_2 \models B} \quad \text{– Modus Ponens}$$

$$\frac{\Gamma \models B}{\Gamma \setminus \{A\} \models A \supset B} \quad \text{– Deduction Rule}$$

We write $A_1, \ldots, A_n \vdash^{d\mathcal{F}}_{\overline{k}} B$ to indicate that $\{A_1, \ldots, A_n\} \models B$ has a general deduction Frege proof of $\leq k$ lines.

Deduction Frege systems are quite general since they allow hypotheses to be "opened" and "closed" (i.e., "assumed" and "discharged") in arbitrary order. A more restrictive system is the *nested deduction Frege* proof system which requires the hypotheses to be used in a 'nested' fashion. The nested deduction Frege systems are quite natural since the correspond to the way mathematicians actually reason while carrying out proofs. A second reason the nested deduction Frege system seems quite natural is that we shall prove below that nested deduction Frege proof systems can simulate with linear size proofs both the propositional Gentzen sequent calculus and tree-like general deduction Frege proofs.

The primary feature of the nested deduction Frege proof system is that hypotheses must be closed in reverse order of their opening. And after a hypothesis is closed, any formula proved inside the scope of the hypothesis is now longer available. For example, a nested deduction Frege proof may look like the following:

$\langle A_1 \rangle \models A_1$	Hypothesis A_1 opened
$\vdots$	
$\langle A_1, A_2 \rangle \models A_2$	Hypothesis A_2 opened
$\vdots$	
$\langle A_1, A_2 \rangle \models A_3$	
$\langle A_1 \rangle \models A_2 \supset A_3$	Deduction Rule; A_2 closed
$\vdots$	
$\langle A_1, A_4 \rangle \models A_4$	Hypothesis A_4 opened
$\vdots$	
$\langle A_1, A_4 \rangle \models A_5$	
$\langle A_1 \rangle \models A_4 \supset A_5$	Deduction Rule; A_4 closed
$\vdots$	
$\langle A_1 \rangle \models A_6$	
$\langle \rangle \models A_1 \supset A_6$	Deduction Rule; A_1 closed

A sequent in a nested deduction Frege ($nd\mathcal{F}$) proof is of the form $\Gamma \models A$ where now Γ is a *sequence* of formulas. An $nd\mathcal{F}$ proof is a sequence of sequents $\Gamma_i \models A_i$ $(i = 1, 2, \ldots, n)$ such that Γ_0 is taken to be the empty sequence and, for each i, one of the following holds:

(a) $\Gamma_i = \Gamma_{i-1}$ and A_i is an axiom.

(b) $\Gamma_i = \Gamma_{i-1} * A_i$. This opens an assumption, $*$ denotes concatenating a formula to the end of the sequence.

(c) Γ_{i-1} is $\Gamma_i * B$ and A_i is $B \supset A_{i-1}$. This is the deduction rule.

(d) $\Gamma_i = \Gamma_{i-1}$ and A_i is inferred from A_j and A_k by Modus Ponens where each of $\Gamma_j \models A_j$ and $\Gamma_k \models A_k$ are available to sequent i. We say sequent j is available to sequent i if $j < i$ and for all ℓ, if $j < \ell < i$ then Γ_j is an initial subsequence of Γ_ℓ.†

We write $A_1, \ldots, A_n \vdash^{nd\mathcal{F}}_{k} B$ if $\langle A_1, \ldots, A_n \rangle \models B$ has a nested deduction Frege proof with $\leq k$ sequents.

Nested deduction Frege proofs can be conveniently represented in pictorial form as a column of formulas with vertical bars that represent the opening, closing and availability of assumptions. This is best defined by an example; the fragmentary $nd\mathcal{F}$-proof given above would be pictorially represented as:

$$\begin{array}{l} \left[\begin{array}{l} A_1 \\ \vdots \\ \left[\begin{array}{l} A_2 \\ \vdots \\ A_3 \end{array}\right. \\ A_2 \supset A_3 \\ \left[\begin{array}{l} A_4 \\ \vdots \\ A_5 \end{array}\right. \\ A_4 \supset A_5 \\ \vdots \\ A_6 \end{array}\right. \\ A_1 \supset A_6 \end{array}$$

Nested deduction Frege proofs are conceptually simple and natural and, in practice, seem to simplify the process of discovering proofs. Thus, it is surprising that a Frege proof system can simulate nested deduction Frege proofs with near linear size proofs: this fact is the content of our main theorems below.

2 Summary of Results

In this section we outline our primary results giving fairly sharp bounds on how much the deduction rule can shorten proofs. Most of our results are stated in the form "If proof system X can prove a formula in n lines, then the formula has a Frege proof of $f(n)$ lines". Obviously f depends on the system X.

First recall the usual proof of the deduction theorem (see e.g. Kleene [5]) which establishes the following:

Theorem 1 *(Deduction Theorem) There is a constant c such that if $A \vdash_n B$ then $\vdash_{c \cdot n} A \supset B$.*

(The constant c is equal to 5 in Kleene's system). The bound of $c \cdot n$ is obtained by replacing each formula C occurring in a proof of B from A with the formula $A \supset C$ and then "filling in the gaps" in the resulting proof with a constant number of lines per gap. For axioms, this is easily done since if C is an axiom then $A \supset C$ can be proved in a constant number of lines. For modus ponens inferences, this is done using the fact that for any formulas A, C and D there is a Frege proof of $A \supset D$ from $A \supset C$ and $A \supset (C \supset D)$ with a constant number of lines. If the proof of Theorem 1 is iterated for m hypotheses, then we get the result that if $A_1, \ldots A_m \vdash_n B$ then $\vdash_{c^m n} A_1 \supset (A_2 \supset \cdots \supset (A_m \supset B) \cdots)$. However we can substantially improve the bound $c^m n$:

Theorem 2 *(Simple Deduction Theorem) Suppose $A_1, \ldots A_m \vdash_n B$. Then*

$$\vdash_{O(n+m)} (A_1 \supset (A_2 \supset \cdots \supset (A_m \supset B) \cdots)).$$

Proof Given an n line proof P of B from assumptions $A_1, \ldots, A_m$, we construct a Frege proof P' of B from the single assumption $A_1 \land A_2 \land \cdots \land A_m$ (where the conjunction is to be associated from left-to-right). For any Frege proof system, there is a constant k such that $C \land D \vdash_k C$ and $C \land D \vdash_k D$. Thus, P' can be constructed to (1) first derive each of $A_1, \ldots, A_m$ in $2k(m-1)$ lines and (2) then derive B in $\leq n$ lines (via P). Clearly P' has $O(m+n)$ lines and by one application of Theorem 1, there is a Frege proof of $A_1 \land \cdots \land A_m \supset B$ with $O(m+n)$ lines. Finally it can be shown by induction on m that

$$\vdash_{O(m)} \quad [A_1 \land \cdots \land A_m \supset B] \supset [(A_1 \supset (A_2 \supset \cdots \supset (A_m \supset B) \cdots))].$$

By combining these last two proofs with a Modus Ponens inference, Theorem 2 is proved. □

We use the name *simple deduction Frege* proof system for the system in which all hypotheses must be opened at the beginning a proof and closed at the end of a proof. Theorem 2 shows that a Frege proof system can simulate simple deduction Frege proofs with linear size proofs; or equivalently, that the simple deduction Frege proof system provides only a *linear speedup* (i.e., a *constant factor speedup*) over Frege proof systems.

† It is also possible, though less elegant, to define sequent j being available to sequent i iff $j < i$ and Γ_j is an initial subsequence of Γ_{i-1}. Our Main Theorems still apply with this alternative definition.

An interesting corollary to Theorem 1 is that conjunctions may be arbirarily reordered and reassociated with linear size Frege proofs:

Corollary 3 *Let B be any conjunction of $A_1, \ldots, A_m$ in that order but associated arbitrarily. Let $i_1, \ldots, i_n$ be any sequence from $\{1, \ldots m\}$ and let C be any conjunction of $A_{i_1}, \ldots, A_{i_n}$ again in the indicated order and associated arbitrarily. Then*

$$\vdash_{\overline{O(m+n)}} B \supset C.$$

Proof By Theorem 1 it suffices to show that $B \vdash_{\overline{O(m+n)}} C$. The proof of B from C proceeds as follows: (1) from the assumption B deduce each subformula of B and, in particular, each of the formulas $A_1, \ldots, A_m$, and (2) deduce each subformula of C from the smallest to the largest. Since there is a constant k such that $E \wedge F \vdash_{\overline{k}} E$ and $E \wedge F \vdash_{\overline{k}} F$ and $E, F \vdash_{\overline{k}} E \wedge F$ for all formulas E and F, it is clear that the proof contains $O(m+n)$ lines. □

We now consider the simulation of proof systems with more powerful versions of the deduction rule.

Theorem 4 *If $\vdash^{d\mathcal{F}}_{\overline{n}} B$ then $\vdash_{\overline{O(n^2)}} B$.*

Theorem 4 states that a general deduction Frege proof system can provide no more than a quadratic speedup over a Frege proof system; whether this quadratic bound is optimal is an open question.

Proof For the proof, we let $\bigwedge_{i=1}^{m} A_i$ denote any conjunction of the formulas A_i ordered and associated arbitrarily (each A_i should occur exactly once as a conjunct). To prove Theorem 4, we prove the more general result that if $\{A_1, \ldots, A_m\} \vDash B$ has a $d\mathcal{F}$-proof P of n lines then $(\bigwedge_{i=1}^{m} A_i) \supset B$ has a Frege proof P' of $O(n^2)$ lines. To form the proof P' replace each sequent $\{A_1, \ldots, A_m\} \vDash B$ of P by the formula $(\bigwedge_{i=1}^{m} A_i) \supset B$; it will suffice to "fill in the gaps" to make P' a valid proof. First, an axiom in P becomes w.l.o.g. an axiom of the Frege system. Second, a hypothesis $\{A\} \vDash A$ in P becomes the tautology $A \supset A$ which has a constant length Frege proof.

Third, the sequents in a Modus Ponens inference in P

$$\frac{\Gamma_1 \vDash A \supset B \qquad \Gamma_2 \vDash A}{\Gamma_1 \cup \Gamma_2 \vDash B}$$

become the formulas $\bigwedge \Gamma_1 \supset (A \supset B)$ and $\bigwedge \Gamma_2 \supset A$ and $\bigwedge(\Gamma_1 \cup \Gamma_2) \supset B$. It will suffice to show that the third formula can be proved from the first formulas with a Frege proof of $O(n)$ lines. By Corollary 3 there are Frege proofs of $\bigwedge(\Gamma_1 \cup \Gamma_2) \supset \Gamma_i$ for $i = 1, 2$ containing $O(m)$ lines where $\Gamma_1 \cup \Gamma_2$ contains m formulas. From these latter two formulas and from $\bigwedge \Gamma_1 \supset (A \supset B)$ and $\bigwedge \Gamma_2 \supset B$ there is a Frege proof of $\bigwedge(\Gamma_1 \cup \Gamma_2) \supset B$ with a constant number of lines. It is easily shown that the number of formulas in the lefthand side of sequent in a $d\mathcal{F}$-proof is bounded by the number of lines in the proof; hence $m \leq n$ and for Modus Ponens, one can "fill in the gap" in P' with $O(n)$ lines.

Fourth, the sequents in a deduction rule inference in P

$$\frac{\Gamma_1 \vDash B}{\Gamma_2 \vDash A \supset B}$$

where Γ_2 is $\Gamma_1 \setminus \{A\}$ become the formulas $\bigwedge \Gamma_1 \supset B$ and $\bigwedge \Gamma_2 \supset (A \supset B)$. In this case, by Corollary 3, there is a Frege proof of $(A \wedge \bigwedge \Gamma_2) \supset \bigwedge \Gamma_1$ of $O(n)$ lines (again since the number of formulas in the conjunction is bounded by n). With this, there is a Frege proof of $\bigwedge \Gamma_2 \supset (A \supset B)$ from $\bigwedge \Gamma_1 \supset B$ with constantly many additional lines. Thus we have "filled the gap" for the deduction rule inference with $O(n)$ lines. □

We next state our main results that Frege systems can simulate nested deduction Frege proof systems with nearly linear proof size. The "near linear" size estimates are in terms of extremely slow growing functions such as *log** and the inverse Ackermann function. The *log** function is defined so that $\log^* n$ is equal to the least number of iterations of the logarithm base 2 which applied to n yield a value < 2. In other words, $\log^* n$ is equal to the least value of k such that $n < 2^{2^{\cdot^{\cdot^{2}}}}$ where there are k 2's in the stack. To get even slower growing functions, we define the $\log^{(*i)}$ functions for each $i \geq 0$. The $log^{(*0)}$ function is just the base 2 logarithm function and the $log^{(*1)}$ is just the log^* function. For $i > 1$, the $log^{(*i)}$ function is defined to be equal to the least number of iterations of the $log^{(*i-1)}$ function which applied to n yields a value < 2. The Ackermann function can be defined by the equations:

$$\begin{aligned} A(0, m) &= 2m \\ A(n+1, 0) &= 1 \\ A(n+1, m+1) &= A(n, A(n+1, m)) \end{aligned}$$

It can be shown that $A(i+1, j)$ is equal to the least value n such that $\log^{(*i)}(n) \geq j$, this means that $\log^{(*i)} A(i+1, j) = j$. It is well-known that the Ackerman function is recursive but dominates eventually every primitive recursive function.

Definition The *inverse Ackerman* function α is defined so that $\alpha(n)$ is equal to the least value of i

such that $A(i,i) > n$. Equivalently, $\alpha(n)$ is equal to the least i such that $\log^{(*i-1)} n < i$.

Main Theorem 5 *Let $i \geq 0$. Suppose $\vdash^{nd\mathcal{F}}_{n} B$ and that in this $nd\mathcal{F}$-proof of B assumptions are opened m times. Then*

$$\vdash_{O(n+m\log^{(*i)} m)} B.$$

Main Theorem 6 *If $\vdash^{nd\mathcal{F}}_{n} B$ then $\vdash_{O(n\cdot\alpha(n))} B$.*

These Main Theorems are extremely close to a linear simulation of nested deduction Frege proof systems by Frege proof systems. Since it is immediate that $m < n$ it follows that if one could somehow bound the number of hypotheses m by $O(n/\log^{(*i)} n)$ for a fixed value i, then one would obtain a linear simulation. However, we have no indication that m can be bounded in this way.

The proofs of the Main Theorems are postponed to the next two sections. First, we discuss and prove some corollaries which give an unexpected connection between nested deduction Frege proof systems and tree-like $d\mathcal{F}$-proofs and the propositional Gentzen sequent calculus. A proof is *treelike* if no line is used more than once in the proof.

Theorem 7 *If $\Gamma \vDash A$ has a treelike general deduction Frege proof of n lines, then $\vdash^{ndF}_{O(n)} \Gamma_{\text{seq}} \vDash A$ where Γ_{seq} is any sequence containing the same elements as the set Γ without repetition.*

Corollary 8 *If A has a treelike $d\mathcal{F}$-proof of n lines, then $\vdash_{O(n\cdot\alpha(n))} A$.*

One elementary fact to note about $nd\mathcal{F}$-proofs is that if Π is a sequence of k formulas and if the sequent $\Pi \vDash A$ has an $nd\mathcal{F}$-proof of n lines, then the sequent also has an $nd\mathcal{F}$-proof of n lines in which the first k lines are hypothesis inferences which open the hypotheses in Π— of course these k hypotheses remain open at the end of the proof. By reordering the first k lines of the $nd\mathcal{F}$-proof, it is clear that for any permutation Π' of Π, $\Pi' \vDash A$ also has an n line $nd\mathcal{F}$-proof. We earlier used the notation $\Pi * A$ to denote the concatenation of a formula A to the end of the sequence Π. For convenience, we define $\Gamma * A$ with Γ a *set* of formulas to be $\Pi_\Gamma * A$ where Π_Γ is any sequence containing all the formulas in Γ (without repetition). This notation will be used only when the order of the formulas in Π_Γ is not important.

To prove Theorem 7 it will suffice to prove the following lemma; since, from an $nd\mathcal{F}$-proof of $\Pi * (\neg B) \vDash p \wedge \neg p$ the sequent $\Pi \vDash \neg B \supset (p \wedge \neg p)$ can be inferred by the deduction rule and from this $\Pi \vDash B$ can be inferred in a constant number of lines. (Here p is an arbitrary propositional variable.)

Lemma 9 *If $\{A_1, \ldots, A_m\} \models B$ has a tree-like general deduction Frege proof of n lines, then $\vdash^{nd\mathcal{F}}_{O(n)} \langle A_1, \ldots, A_m, \neg B\rangle \models p \wedge \neg p$.*

Proof We shall prove by induction on n that, if the sequent $\{A_1, \ldots, A_m\} \models B$ has a tree-like $d\mathcal{F}$-proof P of n lines then there is an $nd\mathcal{F}$-proof P' of $\langle A_1, \ldots, A_m, \neg B\rangle \vDash p \wedge \neg p$ of length $\leq c \cdot n$ lines, for some constant c. The proof splits into four cases depending on the final inference in P.

Case 1: The last line of P is $\models A$, for A an axiom. Then P' is just an $nd\mathcal{F}$-proof of $\langle \neg A\rangle \models p \wedge \neg p$ which has a constant number of lines, say c_1 lines.

Case 2: The last line of P is $\{A\} \models A$. Then P' is an $nd\mathcal{F}$-proof of $\langle A, \neg A\rangle \models p \wedge \neg p$ which has a constant number of lines, say c_2 lines.

Case 3: The last line of P is

$$\frac{\Gamma_1 \models A \supset B \qquad \Gamma_2 \models A}{\Gamma_1 \cup \Gamma_2 \models B}$$

Assume the proof of $\Gamma_1 \models A \supset B$ has n_1 lines and the proof of $\Gamma_2 \models A$ has n_2 lines, so that $n = n_1 + n_2 + 1$ since P is treelike. By the induction hypothesis, there are $nd\mathcal{F}$-proofs P_1 and P_2 of the sequents $\Gamma_1 * (\neg(A \supset B)) \models p \wedge \neg p$ and $\Gamma_2 * (\neg A) \models p \wedge \neg p$ of lengths $\leq c \cdot n_1$ and $\leq c \cdot n_2$ lines, respectively. The proof P' of $\Gamma_1 \cup \Gamma_2 * (\neg B) \models p \wedge \neg p$ is:

$$
\begin{array}{|l}
\Gamma_1 \cup \Gamma_2 \\
\begin{array}{|l}
\neg B \\
\left.\begin{bmatrix} \neg(A \supset B) \\ \vdots \\ p \wedge \neg p \end{bmatrix}\right\} \qquad \text{from } P_1 \\
\neg(A \supset B) \supset (p \wedge \neg p) \\
\vdots \\
A \supset B \\
\left.\begin{bmatrix} \neg A \\ \vdots \\ p \wedge \neg p \end{bmatrix}\right\} \qquad \text{from } P_2 \\
\neg A \supset p \wedge \neg p \\
\vdots \\
A \\
B \qquad\qquad \text{by modus ponens}
\end{array}
\end{array}
$$

This proof has $\leq c \cdot n_1 + c \cdot n_2 + d$ lines where d is a constant. Taking $c \geq d$, the proof has $\leq c \cdot n$ lines.

Case 4: The last line of P is:

$$\frac{\Gamma \models C}{\Gamma \setminus \{A\} \models A \supset C}$$

By the induction hypothesis, there is a $nd\mathcal{F}$-proof P_1 with $c(n-1)$ lines of

$$\Gamma * \neg C \models p \wedge \neg p.$$

The proof of $(\Gamma \setminus \{A\}) * \neg(A \supset C) \models p \wedge \neg p$ is:

$$\left[\begin{array}{l} \Gamma \setminus \{A\} \\ \left[\begin{array}{l} \neg(A \supset C) \\ \quad\vdots \\ \quad A \\ \quad \neg C \\ \left.\begin{array}{l} \quad\vdots \\ \quad p \wedge \neg p \end{array}\right\} \text{ from } P_1 \end{array}\right. \end{array}\right.$$

This proof has size $\leq c(n-1) + d'$ where d' is a constant. So taking $c \geq d'$, the proof has size $\leq c \cdot n$.

Lemma 9 follows from cases 1-4, by taking $c \geq c_1, c_2, d, d'$ and Theorem 7 is proved. □

The next theorem gives a linear simulation of the propositional Gentzen sequent calculus by the nested deduction Frege system. For this theorem, it is crucial that Gentzen sequent calculus proofs are always treelike. For the definition of the Gentzen sequent calculus, see Takeuti [13]; we are concerned with only the propositional fragment of the sequent calculus. The following theorems also hold for many variations of the sequent calculus, for instance with the mix rule, or with a rule that allows arbitrary reordering of cedents, or with either the multiplicative or additive versions of rules. (But the treelike property is crucial for our proofs.)

Theorem 10 *Suppose the sequent* $\Gamma \longrightarrow \Delta$ *has a Gentzen sequent calulus proof of length n lines. Then* $\vdash^{ndF}_{O(n)} \bigwedge \Gamma \supset \bigvee \Delta$.

Corollary 11 *If* $\longrightarrow A$ *has a Gentzen sequent calculus proof of length n lines, then* $\vdash_{O(n \cdot \alpha(n))} A$.

Theorem 10 is proved by showing by induction on n that, if $A_1, \ldots, A_k \longrightarrow B_1, \ldots, B_\ell$ has a Gentzen sequent calculus proof of n lines, then $\vdash^{ndF}_{O(n)} \langle A_1, \ldots, A_k, \neg B_1, \ldots \neg B_\ell \rangle \models p \vee \neg p$. Because of space constraints, the proof of Theorem 10 is omitted from this extended abstract.

Corollary 11 improves a theorem of Orevkov [10] which states that if $\longrightarrow A$ has a proof in a sequent calculus KGI of n lines and height h, then $\vdash_{O(n \log h)} A$. The proof system KGI is a reformulation of the usual sequent calculus [9]; although KGI proofs need not be tree-like, Gentzen proofs must be tree-like in order to be linearly translated in KGI. Orevkov, like us, does not need to count structural inferences.

3 Proof of the Main Theorems

In this section we reduce the main theorems to a serial transitive closure problem which will be considered in the next section.

Suppose that we have a natural deduction Frege proof P of a sequent $\models B$ such that P contains n lines and uses the hypothesis rule m times. To prove Main Theorem 5 for a fixed value of i, we will translate P into a Frege proof of B containing $O(n + m \log^{(*i)} m)$ lines. Likewise, for Main Theorem 6, P is translated into a Frege proof of B of $O(n \cdot \alpha(n))$ lines.

Each line in the proof P is of the form $\Gamma \models B$ where Γ is a sequence of formulas $\langle A_1, \ldots, A_k \rangle$. From the sequent $\Gamma \models B$ we form the logically equivalent formula $(\bigwedge \Gamma) \supset B$ where conjunction is associated from left to right; thus $\bigwedge \Gamma$ is the formula $((\cdots(A_1 \wedge A_2) \wedge \cdots \wedge A_{k-1}) \wedge A_k)$. When Γ is empty, $\bigwedge \Gamma$ is a fixed tautology. This translation of sequents into equivalent formulas gives us a sequence of formulas P'; unfortunately, P' is not a valid Frege proof and so it remains to show how P' can be made into a valid Frege proof with only a relatively small increase in the number of lines.

To make P' into a valid Frege proof we shall add additional lines. There are four rules of inference for $nd\mathcal{F}$: Axiom, Hypothesis, Deduction Rule and Modus Ponens. For each rule of inference, we explain what lines need to be added to P'; we save Modus Ponens for last since it is by far the most difficult case.

First consider an axiom inference in P which is of the form $\Gamma \models B$ where B is an axiom, so P' contains $\bigwedge \Gamma \supset B$ as the corresponding formula. Since a Frege proof system is axiomatized with axiom schemata, there is a proof of $B \supset (X \supset B)$ with a constant number of lines (independent of the formulas B and X). Thus there is a constant length Frege proof of the formula $\bigwedge \Gamma \supset B$; namely, take the axiom B, derive $B \supset (\bigwedge \Gamma \supset B)$ and then use modus ponens. This constant length Frege proof is inserted into the sequence P'.

Second, consider a hypothesis inference where P contains a sequent $\Gamma * B \models B$ and P' contains $((\wedge\Gamma) \wedge B) \supset B$. It is easy to derive $(X \wedge B) \supset B$ in a Frege proof in a constant number of lines where the constant is independent of the formulas B and X, so this sequent in P' can be derived in a constant number of lines.

Third, consider a deduction rule inference in P; here

P contains a sequent $\Gamma * A \models B$ followed immediately by $\Gamma \models A \supset B$ and P' contains the corresponding formulas $((\wedge\Gamma) \wedge A) \supset B$ and $(\wedge\Gamma) \supset (A \supset B)$. Again there is a constant length Frege proof of the latter formula in P' from the former one; this constant length proof is to be inserted into P'.

Fourth and hardest, we consider a line in P' that corresponds to a sequent of P obtained by Modus Ponens. Suppose that in P there are lines $\Gamma_1 \models A$ and $\Gamma_2 \models A \supset B$ from which $\Gamma \models B$ is inferred by Modus Ponens. Since P is a nested deduction Frege proof, Γ_1 and Γ_2 are initial subsequences of Γ. In P' the formulas $(\bigwedge\Gamma_1) \supset A$ and $(\bigwedge\Gamma_2) \supset (A \supset B)$ appear and from them we wish to derive the formula $(\bigwedge\Gamma) \supset B$ in a small number of lines. Note that there is a constant size Frege proof of $X \supset B$ from the hypotheses $X_1 \supset A$ and $X_2 \supset (A \supset B)$ and $X \supset X_1$ and $X \supset X_2$ where the constant is independent of the formulas X, X_1, X_2, A and B. Thus we will modify P' by adding the formulas $(\bigwedge\Gamma) \supset (\bigwedge\Gamma_i)$ at the beginning and inserting a Frege proof of $(\bigwedge\Gamma) \supset B$ from these new formulas and from the other two formulas.

It remains now to give short Frege proofs of the formulas $(\bigwedge\Gamma) \supset (\bigwedge\Gamma')$ which have been added to beginning of P'. By examining the fourth case above we see that there are $< 2n$ such formulas and they always have Γ' an initial subsequence of Γ and thus they are tautologies of the form

$$((\cdots(A_1 \wedge A_2)\wedge\cdots)\wedge A_k) \supset ((\cdots(A_1 \wedge A_2)\wedge\cdots)\wedge A_\ell)$$

where without loss of generality $\ell < k \leq m$. To prove one such tautology requires $O(k-\ell)$ lines. Unfortunately, if we used $O(k-\ell)$ lines for each tautology, the total number of lines would only be bounded by $O(m \cdot n)$ instead of the desired bound of $O(n + m\log^{(*i)} m)$ or $O(n \cdot \alpha(n))$. To get this lower bound on the number of lines we must exploit the fact that there are many tautologies to be proved. In other words, we can achieve significant reduction in the number of proof lines by proving the $2n$ many tautologies simultaneously rather than separately.

To give the short Frege proofs, we shall rephrase the problem as a transitive closure problem. We shall work now only with tautologies of the form $\bigwedge\Gamma \supset \bigwedge\Pi$ where Π is a proper initial subsequence of Γ and may be the empty sequence. Since there were m uses of the hypothesis rule in P, there are $m+1$ distinct $\bigwedge\Gamma$'s; we think of them forming a directed graph G with an edge from $\bigwedge\Gamma$ to $\bigwedge\Pi$ iff Γ extends Π by one element. There are also $\leq 2n$ distinct "target" tautologies which we need to prove. The Frege proof of these target tautologies will proceed as follows: First prove in $O(m)$ lines the tautologies $\bigwedge\Gamma \supset \bigwedge\Pi$ where Γ extends Π by a single element (this generally includes non-target tautologies). Next we prove all the target tautologies in $O(n+m\log^{(*i)} m)$ or $O(n \cdot \alpha(n))$ lines. The procedure for this latter step is to prove many intermediate formulas $\bigwedge\Gamma \supset \bigwedge\Pi$ from the transitive closure of the directed graph of $\bigwedge\Gamma$'s. The details are sketched in the next section.

As an example, for the nested deduction proof pictured in section 1, the directed graph of tautologies is:

$$\begin{tikzcd}[column sep=small] & & \bigwedge\langle A_1, A_2\rangle \arrow[dl] \\ \bigwedge\langle\rangle & \bigwedge\langle A_1\rangle \arrow[l] & \\ & & \bigwedge\langle A_1, A_4\rangle \arrow[ul] \end{tikzcd}$$

4 Serial Transitive Closure Problem

The serial transitive closure problem is the problem of deriving a set of "closure edges" in the transitive closure of a directed graph. If X and Y are nodes in a directed graph we write $X \to Y$ to indicate the presence of an edge from X to Y. Of course, any edge in the transitive closure of a graph can be obtained by a series of *closure steps* which are inferences of the form

$$\frac{A \to B \qquad B \to C}{A \to C}$$

Identifying $\to$ and $\supset$ it is obvious that a closure step can be simulated with a constant number of inferences in a Frege proof.

Serial Transitive Closure of Graphs:

An *instance* consists of (1) a directed graph G with m edges and (2) a list of n *closure* edges $X_i \to Y_i$ $(i = 1, \ldots, n)$ in the transitive closure of G.

A *solution* is a sequence of edges $U_i \to V_i$ $(i = 1, \ldots, s)$ containing all n closure edges such that each $U_i \to V_i$ is inferred by a single closure step from earlier edges and/or edges in G. We call s the number of steps of the solution.

Note that the number of steps in a solution counts only closure steps and does not count edges that are already in G. A directed graph is a *tree* if it is a tree in the usual sense, with root at the top and with all edges downwards (or all edges upwards).

Theorem 12 *Let $i \geq 0$. If the directed graph G is a tree then the serial transitive closure problem has a solution with $O(n + m\log^{(*i)} m)$ steps.*

Theorem 13 *If the directed graph G is a tree then the serial transitive closure problem has a solution with $O((n+m) \cdot \alpha(m))$ steps.*

Theorems 12 and 13 are precisely what is needed to complete the proof of the Main Theorems. This is because in section 3 the proof of the Main Theorems was reduced to the problem of proving $\leq 2n$ 'target' tautologies. Let G be the graph with m edges defined at the end of section 3 and let the target tautologies be the closure edges; then any solution of the serial transitive closure problem leads to a Frege proof of the target tautologies of length $O(s)$.

For the proofs of the two theorems we may assume without loss of generality that G is a rooted tree with edges pointing away from the root. We always picture trees with the root at the top, except in the special case of one-trees, which have fanout 1, the root is to the left and edges point to the right. The concepts of *child, father, ancestor* and *descendent* are defined as usual. The *size* of a tree is defined to be the number of edges in the tree (not the number of nodes). If a tree has e edges, then it has exactly $e+1$ nodes. A *subtree* of a tree is a connected subset of the tree. A subtree is *unscarred* if it is a subtree in the usual sense, i.e., consists of all the edges below a given node in the tree. A subtree S may also be obtained by first removing some set of subtrees and then letting S consist of all the remaining edges below some given remaining node: S is said to have a *scar* at any of its leaf nodes which are roots of earlier removed (nontrivial) subtrees. Two subtrees are said to be *disjoint* if they have no edges in common; disjoint subtrees may share a single node since the root of one may be a scar of the other. If X is a node in T, then T_X denotes the subtree of T rooted at X. The *immediate subtrees* of a tree T are the maximal proper subtrees of T, i.e., the trees T_X for X a child of the root of T. The following lemma is well-known: see, e.g., Brent [1].

Lemma 14 *Let $N \geq 0$ and T be a tree with $\geq N$ edges. Then there is a subtree of T which has size $\geq N$ edges such that each of its immediate subtrees has $< N$ edges.*

Lemma 14 is easily proved by taking a minimal subtree with $\geq N$ edges. The next theorem restates the case $i = 0$ of Theorem 12 with fairly tight bounds on the constants. The *log* function is base two.

Theorem 15 *If the directed graph G is a tree then the serial transitive closure problem has a solution with $n + m \cdot \log m$ closure steps.*

Proof We will first derive $\leq m \log m$ edges, called *auxiliary edges*; each auxiliary edge will be obtained with a single closure step. The choice of auxiliary edges is independent of the closure edges; however, from the edges in G and the auxiliary edges each closure edge can be obtained with (at most) one additional closure step.

For illustration purposes, we first prove the theorem for G a one-tree and then do the general case. Although the general case includes the linear case, the proof of the linear case presents the main ideas more clearly.

Linear Case: Assume G is a one-tree; that is, each node except the leaf has a single child. In this case we may assume the nodes of G are named $X_0, \ldots, X_m$ and that the edges of G are just $X_i \rightarrow X_{i+1}$ for $0 \leq i < m$. The auxiliary edges will be derived in rounds, the first round will in essence split G into two subtrees of $m/2$ edges, the second round splits G into four subtress of $m/4$ edges, etc., for a total of $\lceil \log m \rceil - 1$ rounds. The process is illustrated for the case $m = 8$ in Figure 1; the upper edges of Figure 1 are derived in the first round and the lower edges in the second round.

Round 1: $X_{\lfloor m/2 \rfloor}$ is the midpoint of the one-tree G. The auxiliary edges added in round 1 are the edges of the form $X_j \rightarrow X_{\lfloor m/2 \rfloor}$ for $0 \leq j < \lfloor m/2 \rfloor$ and the edges of the form $X_{\lfloor m/2 \rfloor} \rightarrow X_k$ for $m/2 < k \leq m$. There are exactly m such edges and they can derived with m closure steps if we derive them in the right order; namely, letting j range from $\lfloor m/2 \rfloor - 1$ down to 0 and letting k range from $\lfloor m/2 \rfloor + 1$ up to m. (Actually only $m - 2$ closure steps are needed since two of the auxiliary edges are already in G.)

Round 2: Round 1 split G into two halves; the midpoints of these two halves are $X_{\lfloor m/4 \rfloor}$ and $X_{\lfloor 3m/4 \rfloor}$. In round two, auxiliary edges to and from these midpoints are derived. Namely, (1) the edges $X_j \rightarrow X_{\lfloor m/4 \rfloor}$ with $j < \lfloor m/4 \rfloor$, and (2) the edges $X_{\lfloor m/4 \rfloor} \rightarrow X_j$ with $\lfloor m/4 \rfloor < j \leq \lfloor m/2 \rfloor$, and (3) the edges $X_j \rightarrow X_{\lfloor 3m/4 \rfloor}$ with $\lfloor m/2 \rfloor \leq j < \lfloor 3m/4 \rfloor$, and (4) the edges $X_{\lfloor 3m/4 \rfloor} \rightarrow X_j$ with $\lfloor 3m/4 \rfloor < j \leq m$. There are m such edges and by deriving them in the right order each can be obtained with a single closure step. (Again, taking into account duplicate edges, fewer than m closure steps are needed for round 2.)

Round ℓ: For round number ℓ we add auxiliary edges incident on the nodes $X_{\lfloor k \cdot m/2^\ell \rfloor}$ for odd values of k. Specifically, for each odd value $k < 2^\ell$ the following auxiliary edges are derived: (1) the edges $X_j \rightarrow X_{\lfloor k \cdot m/2^\ell \rfloor}$ for $\lfloor (k-1)m/2^\ell \rfloor \leq j < \lfloor k \cdot m/2^\ell \rfloor$ and (2) the edges $X_{\lfloor k \cdot m/2^\ell \rfloor} \rightarrow X_j$ for $\lfloor k \cdot m/2^\ell \rfloor < j \leq \lfloor (k+1)m/2^\ell \rfloor$. Again, there are exactly m such edges (some of them duplicates of edges from G and edges from earlier rounds); this is seen by using the obvious one-to-one correspondence with the edges of G. And by deriving them in the right order, each auxiliary edge is obtained with a single closure step.

Since there are no more than $\log m$ rounds and fewer than m closure steps are needed in each round, it is

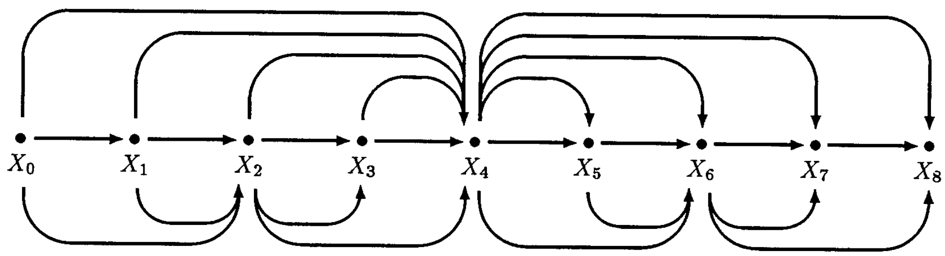

Figure 1

clear that there are $\leq m\log m$ auxiliary edges and that the number of closure steps so far is bounded by $m\log m$.

Now we claim that each of the n closure edges can be obtained with at most a single closure step from the $m\log m$ auxiliary edges (of course some of the closure edges may also be auxiliary edges). To prove this, suppose $X_i \to X_j$ is a closure edge; of course, $i+1 < j$ without loss of generality. Find the least value of ℓ such that for some odd k, $i \leq \lfloor k \cdot m/2^\ell \rfloor \leq j$. If either of the inequalities are actually equalities, then $X_i \to X_j$ is an auxiliary edge added in round ℓ and no additional closure step is needed. If both inequalities are strict, then $X_i \to X_{\lfloor k \cdot m/2^\ell \rfloor}$ and $X_{\lfloor k\cdot m/2^\ell \rfloor} \to X_j$ are both auxiliary edges and from these the closure edge $X_i \to X_j$ can be derived with one closure step.

It follows that all the closure and auxiliary edges are derived with fewer than $n+m\log m$ closure steps and Theorem 15 is proved for G a one-tree.

General Case: The proof of Theorem 15 for G a tree uses a construction similar to the proof of the linear case. For the general case, we shall use Lemma 14 to split G into multiple subtrees of size less than half the size of G (one of these is scarred); then we similarly split these subtrees into subtrees of size less than one quarter the size of G, etc. As in the linear case, we derive auxiliary edges of G as we split G into subtrees; this process will be done in $\leq \log m$ rounds.

Round 1: By Lemma 14 there is a node X in G such that G_X has $\geq m/2$ edges, but the immediate subtrees of G_X have size $< m/2$ edges. Let $G_1, \ldots, G_k$ be the immediate subtrees of G_X. Let G_0 be the tree obtained by removing G_X from G; i.e., G_0 is the scarred subtree with root at the root of G and with a single scar at X.

During round 1, the following auxiliary edges are derived: (1) for each ancestor Y of X the edge $Y \to X$ is an auxiliary edge, and (2) for each descendent Y of X the edge $X \to Y$ is an auxiliary edge. By deriving auxiliary edges in the correct order (namely, shorter edges first), only one closure step is needed for each auxiliary edge. There are at most m auxiliary edges and thus fewer than m closure steps are needed in round 1.

Round 2: Round 1 split G into subtrees $G_0, \ldots, G_k$ of size $\leq m/2$ edges. In round 2 we consider each subtree G_i separately and process it in the manner of round 1. Specifically, suppose G_i has m_i edges; then by Lemma 14, there is a node X in G_i such that $(G_i)_X$ has $\geq m_i/2$ edges and each of its immediate subtrees (in G_i) have size $< m_i/2$. Now auxiliary edges are added from each ancestor of X in G_i to X and from X to each of its descendents in G_i; there are $\leq m_i$ such auxiliary edges and each can be added with at most one closure step. G_i has been split into the following subtrees of size at most $m_i/2$: the immediate subtrees of $(G_i)_X$ and the subtree obtained by removing $(G_i)_X$ from G_i. These subtrees will be treated in the next round.

The total number of closure steps used to derive auxiliary edges in round 2 is less than $m = \sum m_i$.

Round ℓ: The previous round $\ell - 1$ resulted in G being split into multiple, disjoint subtrees of size $\leq m/2^{\ell-1}$. For each such subtree H of size $m_H \geq 2$, Lemma 14 gives a node X in H such that H_X has size $\geq m_H/2$ and each of H_X's immediate subtrees have size $< m_H/2$. As before, auxiliary edges from each ancestor of X in H to X and from X to each descendent of X in H are added in fewer than m_H closure steps. And the immediate subtrees of H_X and the subtree H with H_X removed have size $\leq m_H/2$ and will be treated in next round.

Since the total size of all the disjoint subtrees is equal to m edges, fewer than m closure steps are used in this round.

The process of adding auxiliary edges ends when all the subtrees being considered have size < 2; namely after no more than $\lfloor \log m \rfloor$ rounds. Thus at most

$m \log m$ closure steps are needed for deriving auxiliary edges.

As in the linear case, each of the n closure edges can be obtained with at most a single closure step from the $m \log m$ auxiliary edges. To prove this, suppose Y and Z are nodes in G and $Y \to Z$ is a closure edge; of course, Y is an ancestor of Z. Find the greatest value of ℓ, such that Y and Z are in the same subtree H considered during round ℓ. Unless $Y \to Z$ is already an edge in G, the nodes Y and Z are in different subtrees in the next round. Hence the node X chosen to split subtree H in round ℓ has Y as an ancestor and Z as a descendent. Thus the edges $Y \to X$ and $X \to Z$ are auxiliary edges derived during round ℓ and the closure edge can be added with a single further closure step.

Thus the total number of inference steps needed for a solution of the serial transitive closure problem is less than $n + m \log m$ and Theorem 15 is proved. □

The rest of proof of Theorem 12 proceeds by proving the following theorem by induction on i:

Theorem 16 *Let $i \geq 0$. If the directed graph G is a tree then the serial transitive closure problem has a solution with $(1+2i)(n + m \log^{(*i)} m)$ steps.*

Proof When $i = 0$, the theorem is just a restatement of Theorem 15. So fix $i \geq 1$ and assume the theorem holds for $i-1$. We shall prove the theorem by splitting G into subtrees of size $\log^{(*i-1)} m$, adding auxiliary edges and using the auxiliary edges to derive some of the closure edges; we shall iterate this process $\log^{(*i)} m$ many times after which the subtrees all have size ≤ 1. The derivation of the closure edges from the auxiliary edges will depend on the induction hypothesis.

Let us begin by describing the reduction process which will be used iteratively. The input to the reduction process is a subtree T of G; we assume T has $M > 1$ edges. The output of the reduction process will consist of a set of node-disjoint subtrees of T and the derivation of the closure edges whose endpoints are in T but are in different subtrees output by T. The reduction process has three steps:

Step 1: In the first step, T is partitioned into subtrees and auxiliary edges are derived:

By iteratively applying Lemma 14, T can be split into a finite set of subtrees $T_0, \ldots, T_k$ so that (1) the edges of the T_i's partition the edges of T, and (2) for each $i > 0$, T_i has size $\geq \log^{(*i-1)} M$ edges and (3) for all $i \geq 0$, each immediate subtree of T_i has $< \log^{(*i-1)} M$ edges, and (4) T_0 has root at at the root of T and the rest of the T_i's have a root which is a scar of another subtree in the partition. Clearly there will be at most $\lceil M/\log^{(*i-1)} M \rceil$ many subtrees in the partition.

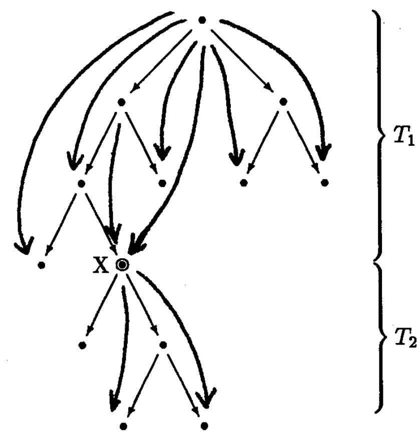

Figure 2

Node X is the root of T_2 and a scar of T_1. Handdrawn edges are the auxiliary edges derived in Step 1. The auxiliary edges of the second kind are the two auxiliary edges with head X; one of these is also of the first kind.

The following auxiliary edges are derived in Step 1: (1) for each T_i with root X, the edges $X \to Z$ for all other nodes Z of T_i are *auxiliary edges of the first kind*; and (2) for each i and j such that the root Y of T_j is a scar of T_i, the edges $X \to Y$ for all ancestors X of Y in T_i are *auxiliary edges of the second kind.* Figure 2 illustrates the choice of auxiliary edges. It is easy to see that by deriving shorter edges first, each auxiliary edge can be derived by single closure step. Further, we claim that there are $\leq 2M$ auxiliary edges. It is easy to see that there $\leq M$ auxiliary edges of the first kind, since each node in T is at the head of at most one such auxiliary edge. To bound the edges of the second kind, note that if T_i has the root Y of T_j as a scar then the ancestors of Y in T_i consist of the root of T_i and some of the nodes in one of the immediate subtrees of T_i. The edge from the root of T_i to Y is also an edge of the first kind and has already been derived. Hence there are $< \log^{(*i-1)} M$ auxiliary edges from nodes inside T_i to Y. Also, the root of T_j can not be the root of T (i.e., $j \neq 0$) so there are at most $M/\log^{(*i-1)} M$ different trees T_j to consider. Taking the product of the number of subtrees and the number of edges, we have that there are less than M auxiliary edges of the second kind.

Step 2: In this step we merely describe the output of the reduction process. The *output trees* are precisely the set of immediate subtrees of $T_0, \ldots, T_k$. Note that the output trees are disjoint and partition the non-root nodes of T but do not contain all the edges of T. Each output tree has $< \log^{(*i-1)} M$ edges.

Step 3: In the third step we derive every closure edge $X \to Y$ with X and Y in T but in different output trees. Let N be the number of such closure edges. We derive these closure edges by setting up a new instance of the serial transitive closure problem and applying the induction hypothesis. The new instance will consist of a directed graph G' which has as nodes the roots of the trees $T_0, \ldots, T_k$ and has as edges the auxiliary edges from Step 1 which connect these roots. The closure edges of the new instance are the edges $X' \to Y'$ which are obtained by the following method: for each closure edge $X \to Y$ (of the original problem) such that X is a node in T_i and Y is a node in T_j with $i \neq j$, let Y' be the root of T_j and let X' be the (scarred) leaf of T_i such that Y is a descendent of X'. It will be important that $X \to X'$ and $Y' \to Y$ are auxiliary edges (of the second and first kind respectively).

Clearly the new instance of the serial transitive closure problem has $< M/\log^{(*i-1)} M$ edges in G' and $\leq N$ closure edges. By the induction hypothesis, it has a solution of size less than or equal to

$$(1+2(i-1))\left[N + \frac{M}{\log^{(*i-1)} M} \log^{(*i-1)}\left(\frac{M}{\log^{(*i-1)} M}\right)\right]$$

which is trivially bounded by

$$(1+2(i-1)) \cdot [N+M].$$

Given a solution to the new serial transitive closure problem, for all X, X', Y and Y' as above, we can derive the closure edge $X \to Y$ in two closure steps from the auxiliary edges $X \to X'$ and $Y' \to Y$ and the closure edge $X' \to Y'$ of the new problem.

To conclude the description of the reduction process, we note that the total number of closure steps needed in the reduction process is bounded by

$$2M + (1+2(i-1))(N+M) + 2N,$$

which is more suggestively written as

$$(1+2i)(N+M).$$

The overall procedure for proving Theorem 16 can now be very simply explained in terms of iterating the above reduction process:

Round 1: Apply the reduction process to the whole tree G. This derives n_1 closure edges (n_1 is the value of N from the reduction process) and outputs a set of subtrees of G which partition the non-root nodes of G and are each of size $< \log^{(*i-1)} m$ edges. The total number of closure steps in round 1 is bounded by

$$(1+2i)(n_1+m).$$

Round ℓ: The previous round generated a set of node-disjoint subtrees each of size less than

$$\underbrace{\log^{(*i-1)}(\log^{(*i-1)}(\cdots(\log^{(*i-1)}}_{\ell-1 \text{ times}}(m))\cdots)).$$

Apply the reduction process (steps 1-3) to all of these subtrees which contain more than one edge; the overall result is that some number n_ℓ of closure edges are derived and that a set of node-disjoint output trees each of size less than

$$\underbrace{\log^{(*i-1)}(\log^{(*i-1)}(\cdots(\log^{(*i-1)}}_{\ell \text{ times}}(m))\cdots))$$

is generated. The total number of closure steps in round ℓ is less than

$$(1+2i)(n_\ell+m).$$

The rounds are iterated until all the subtrees have size ≤ 1; namely, in no more than $\log^{(*i)} m$ rounds. At the end every closure edge has been derived. The total number of closure steps used is bounded by

$$\sum_{\ell=1}^{\log^{(*i)} m} (1+2i)(n_\ell+m)$$

and since $\sum n_\ell \leq n$, the total number of closure edges is bounded by

$$(1+2i)(n+m\log^{(*i)} m).$$

That completes the proofs of Theorems 16, 12 and 5. □

It remains to prove Theorem 13. To do this we first note the following simple corollary of Theorem 16:

Lemma 17 *If the directed graph G is a tree, then the serial transitive closure problem has a solution of size $O(n \cdot \alpha(m) + m \cdot (\alpha(m))^2)$.*

Proof The tree G has m edges. Let $i = \alpha(m)$; this means that $\log^{(*i)} m \leq i$ (in fact, $\log^{(*i-1)} m \leq i$). By Theorem 16, the serial transitive closure problem for G has a solution of size bounded by

$$(1+2\alpha(m)) \cdot (n + m \cdot \alpha(m)). \quad \square$$

As slow-growing functions go, $(\alpha(m))^2$ is not so different from $\alpha(m)$; nonetheless we show how to replace the $\alpha(m)^2$ factor in Lemma 17 by $\alpha(m)$. For this, we shall apply the reduction process used in the proof of Theorem 16. As in step 1 before, G is split into disjoint subtrees; now each contains at least $\alpha(m)$ edges and each one has all immediate subtrees of size less than $\alpha(m)$. Auxiliary edges are derived exactly as in step 1 before. As in step 3, the closure edges which span two different subtrees are derived; by Lemma 17 and the type of counting used in step 3, this takes at most

$$2 \cdot m + c \left[N\alpha\left(\frac{m}{\alpha(m)}\right) + \frac{m}{\alpha(m)}\left(\alpha\left(\frac{m}{\alpha(m)}\right)\right)^2 \right] + 2 \cdot N$$

closure steps where N is the number of closure edges derived in step 3 and c is the constant from Lemma 17. Since $\alpha\left(\frac{m}{\alpha(m)}\right) < \alpha(m)$, the number of closure steps is bounded by

$$2 \cdot m + c \cdot (N \cdot \alpha(m) + m \cdot \alpha(m)) + 2 \cdot N.$$

Instead of iterating the reduction process, we instead just directly derive the rest of the closure edges. If there are remaining N' closure edges then a total of $N' \cdot \alpha(m)$ closure steps suffice; this is because each remaining closure edge connects two nodes inside a subtree of size $< \alpha(m)$.

Since $n = N + N'$, the serial transitive closure problem has a solution of size bounded by

$$(c+2)(n+m) \cdot \alpha(m).$$

Q.E.D. Theorems 13 and 6.

We do not know if the theorems above provide the best possible asymptotic bounds for the serial transitive closure problem or on the size of Frege proof simulations of nested deduction Frege proof systems.

References

[1] R. P. Brent, *The parallel evaluation of general arithmetic expressions*, J. Assoc. Comput. Mach., 21 (1974), pp. 201–206.

[2] S. R. Buss, *The undecidability of k-provability*. to appear in *Annals of Pure and Applied Logic*.

[3] S. A. Cook and R. A. Reckhow, *The relative efficiency of propositional proof systems*, Journal of Symbolic Logic, 44 (1979), pp. 36–50.

[4] K. Gödel, *Über die Länge von Beweisen*, Ergebnisse eines Mathematischen Kolloquiums, (1936), pp. 23–24. English translation in *Kurt Gödel: Collected Works, Volume 1, pages 394-399, Oxford University Press, 1986.*

[5] S. C. Kleene, *Introduction to Metamathematics*, Wolters-Noordhoff and North-Holland, 1971.

[6] J. Krajíček, *Generalizations of proofs*, in Proc. Fifth Easter Conference on Model Theory, Seminarberichte #93, Humboldt Universität, Berlin, 1987, pp. 82–99.

[7] ——, *On the number of steps in proofs*, Annals of Pure and Applied Logic, 41 (1989), pp. 153–178.

[8] J. Krajíček and P. Pudlák, *The number of proof lines and the size of proofs in first-order logic*, Archive for Mathematical Logic, 27 (1988), pp. 69–84.

[9] V. P. Orevkov, *Upper bound on the lengthening of proofs by cut elimination*, Journal of Soviet Mathematics, 34 (1986), pp. 1810–1819. Original Russin version in Zap. Nauchn. Sem. L.O.M.I. Steklov **137**(1984)87-98.

[10] ——, *Reconstruction of a proof from its scheme*, Soviet Mathematics Doklady, 35 (1987), pp. 326–329. Original Russian version in Dokl. Akad. Nauk. **293** (1987) 313-316.

[11] R. J. Parikh, *Some results on the lengths of proofs*, Transactions of the American Mathematical Society, 177 (1973), pp. 29–36.

[12] R. Statman, *Complexity of derivations from quantifier-free Horn formulae, mechanical introduction of explicit definitions, and refinement of completeness theorems*, in Logic Colloquium '86, North-Holland, 1977, pp. 505–517.

[13] G. Takeuti, *Proof Theory*, North-Holland, 2nd ed., 1987.