Proof nets for Herbrand's Theorem

RICHARD MCKINLEY
Universität Bern

This paper explores Herbrand's theorem as the source of a natural notion of abstract proof object for classical logic, embodying the "essence" of a sequent calculus proof. We we see how to view a calculus of abstract Herbrand proofs ("Herbrand nets") as an *analytic* proof system with syntactic cut-elimination. Herbrand nets can also be seen as a natural generalization of Miller's expansion tree proofs to a setting including cut. We demonstrate sequentialization of Herbrand nets into a sequent calculus $\mathbf{LK}_H$; each net corresponds to an equivalence class of $\mathbf{LK}_H$ proofs under natural proof transformations. A surprising property of our cut-reduction algorithm is that it is non-confluent, despite not supporting the usual examples of non-confluent reduction in classical logic.

Categories and Subject Descriptors: F4.1 [**Mathematical logic and formal languages**]: Mathematical logic—*Proof theory*

General Terms: Theory

Additional Key Words and Phrases: Classical logic, cut-elimination, Herbrand's theorem, proof-nets

1. INTRODUCTION

This paper is part of a program [Robinson 2003; Führmann and Pym 2006; 2007; Lamarche and Strassburger 2005a; 2005b; Hughes 2006; Bellin et al. 2006] to understand or uncover the "essence" of proofs in classical logic; the mathematical objects represented by syntactic proofs. This problem traces its roots back to Hilbert's omitted 24th problem [Thiele 2001], which was concerned with "develop(ing) a theory of mathematical proof in general". Such a theory exists and is well-understood for intuitionistic logic; it is provided by the Curry-Howard isomorphism and interpretation in cartesian-closed categories [Lambek and Scott 1986]. Understanding the mathematical theory of classical proof in a similar fashion is still an open problem. Proofs in standard calculi, like the sequent calculus, do not satisfy as mathematical objects, because the essence of a proof is hidden by "bureaucracy": proofs can differ by inessential matters such as the order of in which inferences are applied. For this reason, one approach to uncovering the mathematical structure of proofs is to find "abstract proofs" for classical logic, such that two abstract proofs differ only if the arguments they embody are different. One important part of the study of abstract proofs is *cut-elimination*: given an abstract proof of A implies B, and an abstract proof of B implies C, is there an algorithm yielding an abstract

proof of A implies C? Without discussing in detail the background of this problem (we refer interested readers to the references above), we note that a large part of the problem of representing this operation comes from the unrestricted power of weakening in classical sequent calculus: the so-called "Lafont example" (described in the appendices of [Girard et al. 1989]) uses weakening and cut-elimination as an essential ingredient of an argument that there is exactly one classical proof of every theorem. Avoiding this "collapse" is the first hurdle to be overcome in giving an abstract notion of classical proof with cut-elimination.

Attention in these matters has been paid chiefly to the propositional fragment of classical logic, but this paper looks instead at first-order logic, for which a notion of "essence" is already given by one of the fundamental theorems of logic: Herbrand's theorem [Herbrand 1930]. In its simplest form, Herbrand's theorem states that a formula of first-order logic $\exists x.A$, where A is quantifier free, is provable if and only if there exist ground terms $M_1, \dots M_n$ such that

$$\vDash A[x := M_1] \vee \dots \vee A[x := M_n].$$

This simple form of Herbrand's theorem gives a counterpart in classical logic to the *existence property* of intuitionistic logic: a classical proof of an existential statement does not consist of a single witness, but a (multi)set of candidate witnesses, plus a proof that at least one of them is an actual witness. From a given proof of an existential statement we can extract such a multiset of witnesses, and terms of the "essence" of proofs, it is the point of view of this paper that two proofs of an existential statement have the same essential content if and only if they yield the same multiset of witnesses.

It is well known that a more general "Herbrand's theorem" for formulae in prenex normal form follows directly from Gentzen's cut-elimination theorem [Gentzen 1934], or more properly the *midsequent* theorem (see for example [Troelstra and Schwichtenberg 1996]). The midsequent theorem is usually stated in terms of permutability of inference rules, but it can be more succinctly stated as follows:

THEOREM 1.1 MIDSEQUENT THEOREM. *The cut-free sequent system given in Fig. 1 is complete for sequents of prenex formulae.*

(This statement of the midsequent theorem seems to be novel, although a similar sequent system containing weakening occurs in [Heijltjes 2010]) A proof of a prenex formula $q_1 \dots . q_n.B$ in this calculus yields a set of instantiated versions of B whose disjunction is a tautology: thus the completeness of this calculus can be seen, in itself, as a statement of Herbrand's theorem for prenex formulae. Indeed, a proof in $\mathbf{LK}_H$ is, essentially, the same as an Herbrand proof as formulated by Buss [Buss 1995].

It can be argued (see for example [Hetzl et al. 2008]) that all the mathematically interesting information in a proof in first-order logic is contained in the witnesses used to instantiate the existential quantifiers, and that all other information in the proof is irrelevant to that essence. In particular, two proofs differing only by permuting instances of rules have the same essence. In [Miller 1987], *expansion-tree proofs* were introduced as a formalization of this informal notion of essence: a "Compact Representation of Proofs" in which the inessential details regarding the order of application of rules is discarded. In this paper, we take expansion-tree

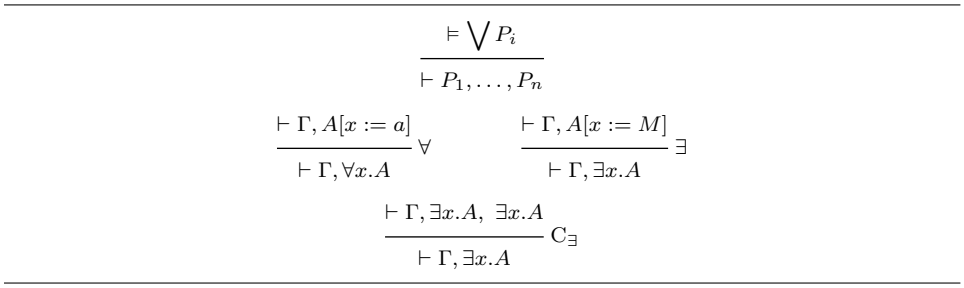


Fig. 1. A "midsequent calculus" $\mathbf{LK}_H^{\alpha\varepsilon}$, sound and complete for prenex classical logic (here the P_i are quantifier-free formulae)

proofs (for first-order logic) and study them as abstract proof objects in the spirit of the program mentioned above.

Classical sequent proofs are very badly behaved under cut-elimination. Cut-elimination is neither confluent nor (and this is more serious) strongly normalizing, and because of this a proof may in general have infinitely many syntactically different normal forms, where normal means cut-free. Without a notion of equality on proofs (which would be given by a good notion of essence) it is difficult to say whether these different normal forms correspond to genuinely different proofs. On the other hand, the typical examples of bad behaviour in Gentzen's system (as detailed in [Girard et al. 1989] and [Girard 1991]) arise where both cut-formulae are the main formula of a structural rule, leading to critical pairs. Observing $\mathbf{LK}_H$, we can see that such an opposition of structural rules cannot occur: weakening is absent, and contraction applies only on existentially quantified formulae. We might hope, therefore, that cut-elimination in the Herbrand setting is better behaved than in the general setting — in particular, we cannot form the Lafont example in $\mathbf{LK}_H$.

We study this question, in this paper, by considering expansion-tree proofs containing cuts, for the restricted case of first-order logic. These proofs with cuts are an example of *proof nets* [Girard 1996], in the sense that they can be studied using the standard toolkit of techniques for dealing with Linear Logic proof nets [Danos and Regnier 1989]. We call this calculus of proof nets Herbrand nets. We show that these nets correspond to proofs in $\mathbf{LK}_H$, giving a correctness criterion for Herbrand nets and a sequentialization theorem. We then develop the theory of cut-elimination inside the Herbrand nets calculus, showing weak normalization, and demonstrate a new counterexample to confluence of cut-reduction which does not rely on the opposition of structural rules in a cut. Since cut-reduction in Herbrand nets lifts to $\mathbf{LK}_H$, the counterexample applies there too, showing that the orientation of critical pairs in classical logic is not enough to guarantee confluence: one must also restrict the permutability of inference steps as in the CBV and CBN fragments of $\bar{\lambda}\mu\tilde{\mu}$ [Curien and Herbelin 2000], and in $\mathbf{LK}^{tq}$ [Danos et al. 1997].

1.1 Related work

Strassburger [Strassburger 2009] has adapted expansion tree proofs to give a notion of proof net for second-order propositional MLL. Proof objects similar to those we present here are also studied in Heijltjes (under the name "Forest proofs")

[Heijltjes 2010], but from a rather different perspective. We will discuss in depth the differences in these two pieces of work later: for now we simply state that our two approaches represent two different ways to repair an intuitive but flawed idea for cut-elimination in expansion-tree proofs. Similar connections between Herbrand's theorem and abstract proof objects for predicate logic were suggested in [Hughes 2006].

2. PRELIMINARY DEFINITIONS

2.1 Prenex formulae of classical first-order logic

A *signature* $\Sigma = (\mathbf{VS}, \mathbf{FS}, \mathbf{PS})$ consists of a countable set **VS** of variable symbols, a countable set **FS** of function symbols, together with a function ar (arity) from **FS** to the natural numbers, and a countable set **PS** of predicate symbols, together with a function Ar from **PS** to the natural numbers. A *constant* of a signature Σ is a function symbol with arity zero. We will use metavariables x, y, z, a, b to denote variable symbols, f, g to denote function symbols, and p, q to denote predicate symbols. The *first-order terms* of Σ are given by the following grammar:

$$M ::= x \mid f(M_1, \dots M_{ar(f)}).$$

Given a term M, the *free variables of* M (written $\mathsf{free}(M)$) are defined as follows:

$$\mathsf{free}(x) = \{x\},$$

$$\mathsf{free}(f(M_1, \dots M_n)) = \mathsf{free}(M_1) \cup \dots \cup \mathsf{free}(M_n).$$

An *atomic formula* is a tuple consisting of a polarity from $\{+, -\}$, a predicate symbol p of arity n, and n terms $M_1, \dots M_n$. We will write an atomic formula $(+, p, M_1, \dots, M_n)$ as $p(M_1, \dots M_n)$, and an atomic formula $(-, q, N_1, \dots, N_n)$ as $\bar{q}(N_1, \dots N_n)$.

The quantifier-free formulae (QFFs) are generated from the atomic formulae using the connectives $\wedge$ and $\vee$:

$$P, Q := p(M_1, \dots M_{\mathrm{Ar}(p)}) \mid \bar{p}(M_1, \dots M_{\mathrm{Ar}(p)}) \mid (P \vee Q) \mid (P \wedge Q)$$

Notice that we give no explicit connective for negation; instead we present formulae in *negation normal form*. Each formula A has a dual formula $\bar{A}$ defined by *De Morgan duality*:

$$\overline{p(M_1, \dots M_n)} = \bar{p}(M_1, \dots M_n) \qquad \overline{\bar{p}(M_1, \dots M_n)} = p(M_1, \dots M_n)$$

$$\overline{(P \vee Q)} := (\bar{P} \wedge \bar{Q}), \quad \overline{P \wedge Q} := \bar{P} \vee \bar{Q}.$$

A formula in prenex normal form (or *prenex formula* for short) is a member of the following grammar, where x ranges over the variables in **VS** and P over QFFs:

$$A ::= P \mid \exists x.A \mid \forall x.A$$

The dual of a prenex formula is defined, as for QFFs, using De Morgan duality:

$$\overline{\forall x.A} := \exists x.\bar{A}, \quad \overline{\exists x.A} := \forall x.\bar{A}$$

The *rank* of a prenex formula is the number of quantifier instances in its prefix. The bound and free variables of a prenex formula are defined as usual: we use the

notation $\mathsf{free}(A)$ and $\mathsf{bound}(A)$ to denote the sets of free and bound variables of a formula A. Notice that, because of the way prenex formulae are built, for any prenex formula A we have $\mathsf{free}(A) \cap \mathsf{bound}(A) = \emptyset$. We will use the notation $A[x := M]$ for the usual notion of substitution of a first-order term M for a variable x in a formula A.

3. EXPANSION TREES AND $\alpha\varepsilon$-FORESTS

As representations of proofs, sequent proofs (for example in $\mathbf{LK}_H$) are unsatisfactory in the sense that they lack *canonicity*. This manifests in the order of application of rules; we can find two proofs of the same formula which differ only by a permutation of two non-interfering rules. Miller's expansion-trees [Miller 1987] provide a better notion of abstract proof, where the linear ordering on quantifier occurrences induced by an $\mathbf{LK}_H$ derivation is replaced by a dependency relation induced by quantifier nesting and variable dependencies. An expansion-tree forms an *expansion-tree proof* of a prenex formula if the dependency relation induced is *irreflexive*: that is, irreflexivity of the dependency relation is a *correctness criterion* for expansion-tree proofs. Expansion-tree proofs provide a form of abstract proof only for *cut-free* proofs, and there is no existing notion of cut-reduction on expansion-tree proofs. In the following section, we give a reformulation of expansion tree proofs (restricted to the case of first-order prenex formulae), extended to account for multiple conclusions and the presence of cuts. We call this extended calculus *Herbrand nets*, since as we will see they are closely related to Girard's proof nets for linear logic. We discuss in the conclusion of the paper the possibility of extending this generalization to the full range of logics captured by expansion-tree proofs (including non-prenex formulae and higher-order quantification). In the presence of cuts, acyclic dependency is not enough to check correctness; in the section following this one, we will use an adapted form of proof-net correctness to identify the correct proofs.

3.1 $\alpha\varepsilon$ terms

In this section we define $\alpha\varepsilon$-terms, which consist of the *expansion-trees* (a reformulation of Miller's expansion trees for the prenex first-order fragment of classical logic), *cuts*, and *witnessing terms*. These trees will form the basis of the Herbrand nets we will define later.

Definition 3.1 *$\alpha\varepsilon$ terms.* Let $\Sigma = (\mathbf{VS}, \mathbf{FS}, \mathbf{PS})$ be a signature, and let $\mathbf{I}$ be a countable set of indices. The *$\alpha\varepsilon$ terms* $t, \ldots$ over $(\Sigma, \mathbf{I})$ (consisting of the *expansion trees* $p, \ldots$, *cuts* $c, \ldots$, and *witnessing terms* $w, \ldots$) are given by the following grammars:

$$t := e \mid w \mid c$$

$$p := S \mid \alpha[a].e \mid (w + \cdots + w)$$

$$w := \varepsilon[M].e$$

$$c := e \bowtie e$$

where S is a nonempty finite set of indices, M is a first-order term over the signature, $a \in \mathbf{VS}$, and $(w + \cdots + w)$ denotes a finite nonempty formal sum (a member of the

free commutative semigroup over w). A *non-cut term* is either an expansion tree or a witnessing term.

Remark 3.2. Expansion-tree proofs were introduced to give a higher-order analogue of Herbrand's theorem (where one cannot rely on Skolem functions or a restriction to formulae in prenex normal form). Why then do we only consider expansion-trees for first-order prenex formulae? Our goal is to find abstract proofs which can be seen as the underlying objects of a sequent calculus, and on which operations such as cut-reduction can be performed directly, without needing to translate back to the sequent calculus. This works for prenex formulae, because there is a strong connection between $\mathbf{LK}_H$ derivations and expansion trees. This strong connection is lost once we move to the setting of full first-order logic: a sequent calculus corresponding to general Herbrand proofs require some *deep* contraction (contraction of existential *subformulae*; this can be seen in Miller's original paper), about which very little can be said in terms of structural proof theory; certainly, syntactic cut-elimination for such a system would be very challenging. For this reason, we concentrate on the prenex fragment in this paper. We give some perspectives on moving beyond that fragment in the conclusions of the paper.

The witnessing terms represent the components of (generalized) Herbrand disjunctions. We make an explicit distinction between the witnessing term $\varepsilon[M].t$ and the expansion tree $(\varepsilon[M].t)$. We will refer to a witnessing term not in the scope of a semigroup $+$ as a *naked* witness.

Remark 3.3. The reader might wonder why we have a commutative semigroup rather than commutative monoid structure on expansion trees: why are we not allowed to form the empty formal sum as a expansion tree? Nontrivial expansions (containing more than one witness) correspond to contraction in the sequent calculus: similarly, allowing empty expansions would amount to explicit weakening in our sequent calculus, and in the proof nets we will form from $\alpha\varepsilon$ terms. Weakening is notoriously difficult to handle well in proof nets; in this setting explicit weakening is not necessary, and we avoid the problems that weakening usually causes for classical proof nets.

3.2 Typing $\alpha\varepsilon$-terms

We now assign *types* to these terms. Note that a typing judgement $t : A$ should not be seen as a proof of A, just as a proof-structure in **MLL** with conclusion Γ is not a proof of Γ. The type of an expansion tree is always a prenex formula. The witnessing terms and cuts receive special non-logical types:

Definition 3.4. A *type* over a signature $\Sigma = (\mathbf{VS}, \mathbf{FS}, \mathbf{PS})$ is either

(a) A *logical type*: a formula of classical predicate logic in prenex normal form over the signature; or

(b) a *non-logical* type, of which there are two kinds:
   i A *witness type*, written $\langle \exists x.A \rangle$, where $\exists x.A$ is a formula in prenex normal form; or
   ii A *cut type*: a pair of dual formulae of classical logic in prenex normal form, written $A \bowtie \bar{A}$.

$$\frac{i_1, \ldots i_n \in \mathbf{I}}{\{i_1, \ldots i_n\} : P}$$

$$\frac{t : A[x := a]}{\alpha[a].t : \forall x.A} \qquad \frac{t : A[x := M]}{\varepsilon[M].t : \langle \exists x.A \rangle}$$

$$\frac{w_1 : \langle \exists x.A \rangle,\ \ldots,\ w_n : \langle \exists x.A \rangle}{(w_1 + \cdots + w_n) : \exists x.A}$$

$$\frac{t : A \quad s : \bar{A}}{t \bowtie s : A \bowtie \bar{A}}$$

Fig. 2. Typing derivations for $\alpha\varepsilon$ terms

We will occasionally need to refer to a type without specifying if it is logical or non-logical: in that case we will use a capital T, reserving $A, B, \ldots$ for those types which are prenex formulae.

We use the witness types to distinguish between a witness, $\varepsilon[M].s$, which receives a witness type, and the expansion tree $(\varepsilon[M].s)$, which receives a logical type. We make this distinction because it will force our proof-nets to have canonical n-ary contractions. Each non-logical type has an underlying logical type:

Definition 3.5. The *underlying type* of a witness type $\langle \exists x.A \rangle$ is $\exists x.A$. The *underlying type* of $A \bowtie \bar{A}$ is A. The free/bound variables free and bound of a witness/cut type are the free/bound variables of its underlying type. We define substitution into witness/cut types in the obvious way

$$\langle \exists x.A \rangle[y := M] = \langle \exists x.A[y := M] \rangle$$

$$(A \bowtie \bar{A})[y := M] = A[y := M] \bowtie \bar{A}[y := M]$$

Definition 3.6. A typed term is a pair $t : T$ of a term t and a type T, derivable in the typing system given in Fig. 2.

There are some terms that cannot be typed, for simple reasons. For example, the term $\alpha[a].t \bowtie \alpha[b].s$ can never be well-typed: a type for a term beginning with an α must be a formula of the form $\forall x.A$, and two such formulae can never be dual.

Example 3.7. *The following is a well-typed term, which will be an important example for us for the rest of the paper. Its type is the drinker's formula("in every bar, there is a patron such that, if she drinks, then everyone drinks"): for that reason we will call it D, the drinker's term:*

$$D = (\varepsilon[\mathsf{c}].\alpha[a].\{1\} \ + \ \varepsilon[a].\alpha[b].\{1\}) : \exists x.\forall y(\bar{A}(x) \vee A(y))$$

The construct $\alpha[a]$ should be thought of as binding a: thus we have the notion of α-bound and α-free variables:

Definition 3.8. Let $t : T$ be a typed term. We define two sets of variables $\mathsf{bound}_\alpha(t : T)$ (the variables α-*bound* in $t : T$) and $\mathsf{free}_\alpha(t : T)$ (the α-free variables of $t : T$) as follows:

(a) The variable a is a member of $\mathsf{bound}_\alpha(t : T)$ if and only if t has a subterm of the form $\alpha[a].s$.

(b) The set $\mathsf{free}_\alpha(t : T)$ is defined as follows:
- $\mathsf{free}_\alpha(S : P) = \mathsf{free}(P)$
- $\mathsf{free}_\alpha(\alpha[a].t : \forall x.A) = \mathsf{free}_\alpha(t : A[x := a]) \setminus \{a\}$
- $\mathsf{free}_\alpha(\varepsilon[M].t : \langle \exists x.B \rangle) = \mathsf{free}_\alpha(t : B[x := M]) \cup \mathsf{free}(M)$
- $\mathsf{free}_\alpha((t_1 + \cdots + t_n) : \exists x.B) = \mathsf{free}_\alpha(t_1 : \langle \exists x.B \rangle) \cup \cdots \cup \mathsf{free}_\alpha(t_n : \langle \exists x.B \rangle)$
- $\mathsf{free}_\alpha(t \bowtie s : A \bowtie \bar{A}) = \mathsf{free}_\alpha(t : A) \cup \mathsf{free}_\alpha(s : \bar{A})$

EXAMPLE 3.9. *For the typed expansion tree $t : A$ below,*

$$t : A = (\varepsilon[b].\alpha[a].(\varepsilon[a].\{1\})) : \exists x.\forall y.\exists z.P(x, y, z, w)$$

if $\{x, y, z, w\}$ is the set of free variables of the QFF $P(x, y, z, w)$, then $\mathsf{free}_\alpha(t : A) = \{b, w\}$ and $\mathsf{bound}_\alpha(t : A) = \{a\}$.

An expansion-tree proof, in the sense of Miller, is a single tree t and proves a single formula A. We will need to extend this idea to *forests* of expansion trees, or more generally, forests of expansion-trees, witnesses and cuts. Such forests of typed terms will play for us the role of *proof-structures*; objects which locally have the structure of a proof, but which might not satisfy our correctness criterion. However, not every forest of typed terms can be regarded as a proof structure: for example, the correctness criterion we define will rely on there being at most one subterm of the form $\alpha[a].t$ for each variable a — that is, we will need a form of *eigenvariable condition*. The following definition pins down our notion of proof-structure, the $\alpha\varepsilon$-*forests*:

Definition 3.10. Let F be a forest built from typed terms.

(a) A variable a is α-*bound in* F *($a \in \mathsf{bound}_\alpha(F)$)* if it is in $\mathsf{bound}_\alpha(t : A)$, for some term $(t : A)$ in F.

(b) The variable a is α-*free in* F ($a \in \mathsf{free}_\alpha(F)$) if it is in $\mathsf{free}_\alpha(t : A)$, for some term $(t : A)$ in F, and not α-bound in F.

(c) F is an $\alpha\varepsilon$-*forest* if
- i each occurrence of $\alpha[a]$ in F is associated with a unique eigenvariable a, and
- ii for each non-cut root $t : A$ of F, $\mathsf{bound}_\alpha(F) \cap \mathsf{free}(A) = \emptyset$.

Each $\alpha\varepsilon$-forest has a type: the multiset consisting of the types of its non-cut roots. Given an $\alpha\varepsilon$-forest, denote by Ind_F the set of tautology indices occurring in F. We consider $\alpha\varepsilon$-forests modulo the renaming of eigenvariables, and also modulo the renaming of tautology indices. We use the notation $[a \leftarrow b]$ to denote the renaming of an α-bound variable, and $[i \leftarrow j]$ for the renaming of an index i.
We use the shorthand $(t : T)[a \leftarrow b]$ for $t[a \leftarrow b] : T[a := b]$ (note that a may only appear in T if T is a cut type; otherwise t and $t[a \leftarrow b]$ have the same

type). Define the renaming of a variable in an $\alpha\varepsilon$-forest pointwise on its roots: if $F = t_1 : T_1, \ldots, t_n : T_n$ is an $\alpha\varepsilon$-forest, then

$$F[a \leftarrow b] := (t_1 : T_1)[a \leftarrow b], \ldots, (t_n : T_n) : [a \leftarrow b]$$

and

$$F[i \leftarrow j] := (t_1 : T_1)[i \leftarrow j], \ldots, (t_n : T_n) : [i \leftarrow j].$$

We will use the following notation for renaming a set of variables/indices occurring in an $\alpha\varepsilon$ forest:

Definition 3.11. Let $V = v^1, \ldots, v^n$ be a set of variable symbols, and $I = i^1, \ldots, i^m$ a set of tautology indices occurring in an $\alpha\varepsilon$ forest. Let $V_i = v^1_j, \ldots, v^n_j$ be sets of variable symbols and $I_j = i^1_j, \ldots, i^m_j$ be sets of indices, for $j \in \{0, 1\}$ such that $V_0 \cap V_1 = \emptyset$, $I_0 \cap I_1 = \emptyset$, and such that no member of V_j or I_j occurs in F. Then define

$$\tau_j(t) := t[v^1 \leftarrow v^1_j] \ldots [v^n \leftarrow v^n_j][i^1 \leftarrow i^1_j] \ldots [i^m \leftarrow i^m_j]$$

Suppose that F is an $\alpha\varepsilon$-forest containing a cut $\alpha[a].t \bowtie (\varepsilon[M].s)$. The intuitive explanation of the cut is a pending communication: during cut-elimination, the witness M, will be substituted everywhere for the eigenvariable a.

Definition 3.12. Let F be a $\alpha\varepsilon$-forest, a a variable with $a \notin \mathsf{bound}_\alpha F$, and M a term with $\mathsf{free}(M) \cap \mathsf{bound}_\alpha(F) = \emptyset$. We define an operation $[a := M]$ (substitute M for a) on $\alpha\varepsilon$-forests F such that $a \notin \mathsf{bound}_\alpha(F)$. On witnessing terms, of the form $\varepsilon[N].t$, the substitution applies inside the instantiating first-order term M and in the remaining subterm t:

$$\varepsilon[N].t\ [a := M] = \varepsilon[N[a := M]].(t[a := M])$$

Substitution is pushed past all the other term constructors, as follows:

$$\begin{aligned} S[a := M] &= S \\ (\alpha[d].t)[a := M] &= \alpha[d].(t[a := M]) \\ (t_1 + \cdots + t_n)[a := M] &= (t_1[a := M] + \cdots + t_n[a := M]) \\ (t \bowtie s)[a := M] &= t[a := M] \bowtie s[a := M] \end{aligned}$$

Finally, $F[a := M]$ is defined as the pointwise substitution of M for a in each term of F.

By induction on the structure of typing derivations, we obtain:

PROPOSITION 3.13. *If t can be assigned type T, then $t[a := M]$ can be assigned type $T[a := M]$.*

4. HERBRAND NETS

The *correctness* problem for a class of proof structures is the problem of providing an algorithm singling out just those structures arising from a sequential derivation – a *correctness criterion*. In our setting, this amounts to giving a function from $\mathbf{LK}_H$ derivations to $\alpha\varepsilon$-forests, and a criterion identifying just those $\alpha\varepsilon$-forests

arising from an $\mathbf{LK}_H$ derivation. In this section, we define such a criterion, and prove it has the *sequentialization* property: from any F satisfying our criterion, we can recover a sequent derivation yielding F. The techniques we use are, in most cases, minor variations on standard techniques for first-order **MLL** without units; where proofs are more than a few lines long, we present them in Appendix A.

4.1 $\alpha\varepsilon$-forests as proof structures

We consider proof structures to be forests with links – a relation on the subtrees of the forest. The links on an **MLL** proof net are simply the axiom links connecting dual atoms. The linking structure on an $\alpha\varepsilon$-forest is given using *jumps* [Girard 1996]. If the variable x appears free in a first-order term M, there is a jump from each $\varepsilon[M]$ to the alpha node binding x. This jump indicates that, in a sequent derivation of F, the existential rule introducing the $\varepsilon[M]$ must occur above the universal rule introducing the $\alpha[a]$ in any sequentialization. Less obviously, we also need jumps from cuts: if the variable a is free in the type of a cut, then that cut must occur above the rule binding a. The usual axiom links of proof nets, linking two dual formulae, are replaced in Herbrand nets by something more general: the information contained at the leaves of an $\alpha\varepsilon$-forest plays the role of generalized axiom links. This generalization is two-fold: each "tautology link" (each index appearing in a set at some leaf) may have an arbitrary (finite) number of conclusions, and (because of contraction) each leaf may be connected to several such links. We also represent this information with jumps, which behave similarly to the quantifier jumps. We will call this graph with jumps the *dependency graph* of the forest.

Definition 4.1. Let F be an $\alpha\varepsilon$-forest with the eigenvariable property. The *dependency graph* $\mathrm{Dep}(F)$ of F is a labelled directed graph whose vertices are:

(a) The occurrences of subterms of F, plus

(b) one *tautology node* for each tautology index $i \in \mathrm{Ind}_F$, labelled with i.

The edges of $\mathrm{Dep}(F)$ are the edges of F considered as a directed graph (with edges directed toward the roots), plus the *jumps*:

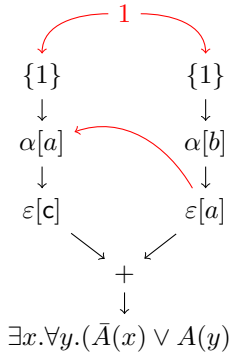
—An edge from $\varepsilon[M].s$ to $\alpha[a].t$ whenever $a \in \mathsf{free}(M)$;

—An edge from $t \bowtie s : A \bowtie \bar{A}$ to $\alpha[a].u$ whenever $a \in \mathsf{free}(A)$

—An edge from the vertex i to each leaf S of F with $i \in S$.

When drawing the dependency graph, we use red curved arrows to represent jumps and red labels for the tautology vertices; the black, straight arrows and black vertices represent the underlying forest structure. We refer to the vertices of the dependency graph as *nodes*. The nodes fall into several families; S is a propositional node, $\alpha[a].t$ an α-node, $\varepsilon[M].t$ an ε-node, and $(w_1 + \cdots + w_n)$ an *expansion* node.

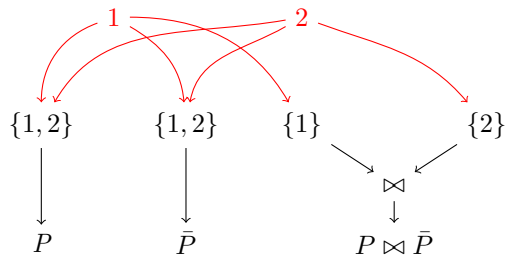
EXAMPLE 4.2. *The dependency graph of the drinker's term D is*



EXAMPLE 4.3. *The dependency graph of the $\alpha\varepsilon$-forest*

$$\{1,2\} : P,\ \{1,2\} : \bar{P},\ P \bowtie \bar{P}$$

is



The dependency graph induces a relation (which we call *dependency*) on the nodes of an $\alpha\varepsilon$-forest: we will write $t \lhd s$ when t and s are subtrees of F and there is a directed path from s to t in the dependency graph of F.

4.2 Correctness

We use a variation on the well-known ACC (ACyclic Connected) criterion [Danos and Regnier 1989] to define correctness. The criterion as given is exponential (we can decide in exponential time if a given $\alpha\varepsilon$-forest is ACC correct), but it is known that correctness for this kind of proof-net is actually NL-complete [de Naurois and Mogbil 2007]. Of course, checking that a given F is an Herbrand net can be much worse than polynomial, depending on the theory over which we work: in particular, if there are no non-logical axioms in our theory then checking correctness is co-NP complete.

The crucial notions in ACC correctness are the switching and the switching graph, which in our setting are defined for strict typed forests (and not just annotated sequents) as follows:

Definition 4.4. Let F be an $\alpha\varepsilon$-forest.

(a) The *switched nodes* of F are the subterms of the form $\alpha[a].t'$, $(t_1 + \cdots + t_n)$, or S. All other nodes of F are unswitched.

$$\frac{\vDash \bigvee_{j=1}^{n} P_j}{\{i\} : P_1, \ldots, \{i\} : P_n}\, i$$

$$\frac{F,\ t : A[x := a]}{F,\ \alpha[a].t : \forall x.A}\,\forall \qquad \frac{F,\ t : A[x := M]}{F,\ (\varepsilon[M].t) : \exists x.A}\,\exists$$

$$\frac{F,\ t : \exists x.A,\ s : \exists x.A}{F,\ t + s : \exists x.A}\,\mathrm{C}_\exists \qquad \frac{F,\ S : P,\ T : P}{F,\ S \cup T : P}\,\mathrm{C}_\mathrm{P}$$

$$\frac{F,\ t : A \quad G,\ s : \bar{A}}{F,\ G,\ t \bowtie s : A \bowtie \bar{A}}\,\textsc{Cut}$$

Fig. 3. $\mathbf{LK}^{\alpha\varepsilon}_H$: An annotated sequent calculus for prenex classical logic

(b) A *switching* σ of F is a choice of, for each switched link t of F, exactly one incoming edge for t in $\mathrm{Dep}(F)$.

(c) The *switching graph* F_σ of a switching σ is the undirected graph derived from $\mathrm{Dep}(F)$ by deleting, for each switched node t, all edges coming into t except that chosen by the switching, and then forgetting directedness of edges.

Definition 4.5. an $\alpha\varepsilon$-forest F is *ACC-correct* (or just ACC), if for each switching σ, F_σ is connected and acyclic.

In addition to checking ACC correctness, we also need to check that the disjunction of the formulae arising from a tautology index is really a tautology:

Definition 4.6. Let F be an $\alpha\varepsilon$-forest, and let i be a tautology index appearing in F. The formula F_i is defined as follows:

$$F_i = \bigvee \{A \mid (S) : A \text{ is a propositional node in F},\ i \in S\}$$

Definition 4.7. An annotated sequent F is an *Herbrand net* if is ACC-correct, has no naked witnesses, and if for each tautology index i in F, we have $\mathcal{T} \vDash F_i$.

Proposition 4.8. *(a) $F, \alpha[a].t : \forall x.A$ is ACC correct iff $F, t : A[x := a]$ is ACC correct and $a \notin \mathsf{free}_\alpha(F)$.*

(b) $F, (w_1 + \cdots w_n) : \exists x.A$ is ACC correct iff $F, w_1 : \langle \exists x.A \rangle, \ldots w_n \langle \exists x.A \rangle$ is ACC correct.

(c) $F, S : P$ is ACC correct iff F is ACC correct.

Proof. An easy application of the definition of correctness; in each case, we add/remove a switched node which is a root. This cannot affect either connectedness or cyclicity of the switching graph. □

4.3 Decorating sequent derivations with terms

To make explicit the connection between sequential proofs and proof nets, we must give a function from sequent proofs to proof nets. We do this by using $\alpha\varepsilon$ terms to decorate the formulae appearing in sequent proofs, similarly to how one may assign lambda terms to proofs of intuitionistic logic. This *annotated* $\mathbf{LK}_H$ is given in Fig. 3. The rules of annotated $\mathbf{LK}_H^{\alpha\varepsilon}$ operate not on sequents, but on $\alpha\varepsilon$-forests whose types are classical sequents. In order to ensure that the conclusion of a sequent proof s an $\alpha\varepsilon$-forest, we must use eigenvariables *strictly*: each instance of the universal quantifier should have a unique associated eigenvariable, and that eigenvariable should appear free only in the subproof above the rule introducing that quantifier. We must also insist that each instance of the tautology rule has a unique index.

Definition 4.9. A derivation in $\mathbf{LK}_H^{\alpha\varepsilon}$ is a tree built from rule instances from Fig. 3, with instances of the tautology rule at the leaves. A derivation Φ is *strict* if

(i) each tautology rule in Φ is labelled with a distinct index i,

(ii) An eigenvariable a does not appear free in the type of any sequent outside the subproof above the rule introducing $\alpha[a]$.

We write $\mathbf{LK}_H^{\alpha\varepsilon} \vdash F$ if there is a *strict* derivation in $\mathbf{LK}_H^{\alpha\varepsilon}$ of F.

Note that case (ii) in the above definition ensures that eigenvariables are used strictly in the usual sense, and additionally enforces the usual variable restriction on the rule for the universal quantifier.

Remark 4.10. The annotated system $\mathbf{LK}_H^{\alpha\varepsilon}$ provides a *canonical* function from $\mathbf{LK}_H$ proofs to $\alpha\varepsilon$-forests (modulo renaming of indices). Such a canonical function does not exist for Robinson's proof nets [Robinson 2003], owing to the presence of weakening; by working in the absence of weakening, we avoid this problem.

Example 4.11. *Let Σ contain the unary predicate A and a constant symbol c. Recall the drinker's term D (Example 3.7):*

$$D = (\varepsilon[\mathsf{c}].\alpha[a].\{1\} + \quad \varepsilon[a].\alpha[b].\{1\}) : \exists x.\forall y(\bar{A}(x) \vee A(y)) \tag{1}$$

D is the conclusion of the derivation below:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\{1\} : \bar{A}(\mathsf{c}) \vee A(a),\quad \{1\} : \bar{A}(a) \vee A(b)}\,1}{\{1\} : \bar{A}(\mathsf{c}) \vee A(a),\quad \alpha[b].\{1\} : \forall y \bar{A}(a) \vee A(y)}\,\forall\mathrm{R}}{\{1\} : \bar{A}(\mathsf{c}) \vee A(a),\quad (\varepsilon[a].\alpha[b].\{1\}) : \exists x.\forall y(\bar{A}(x) \vee A(y))}\,\exists\mathrm{R}}{\alpha[a].\{1\} : \forall y(\bar{A}(\mathsf{c}) \vee A(y)),\quad (\varepsilon[a].\alpha[b].\{1\}) : \exists x.\forall y(\bar{A}(x) \vee A(y))}\,\forall\mathrm{R}}{(\varepsilon[\mathsf{c}]\alpha[a].\{1\}) : \exists x.\forall y(\bar{A}(x) \vee A(y)),\quad (\varepsilon[a].\alpha[b].\{1\}) : \exists x.\forall y(\bar{A}(x) \vee A(y))}\,\exists\mathrm{R}}{(\varepsilon[\mathsf{c}].\alpha[a].\{1\} \ + \ \varepsilon[a].\alpha[b].\{1\}) : \exists x.\forall y(\bar{A}(x) \vee A(y))}\,\mathrm{C}_\exists \tag{2}$$

The following result immediately gives completeness of Herbrand nets with respect to prenex classical logic:

Proposition 4.12. *The conclusion of any $\boldsymbol{LK}_H^{\alpha\varepsilon}$ derivation is an Herbrand net.*

PROOF. By induction on the tree-structure of an $\mathbf{LK}_H^{\alpha\varepsilon}$ proof. □

Two derivations in annotated $\mathbf{LK}_H^{\alpha\varepsilon}$ derive the same Herbrand net if and only if they can be derived from each other by a sequence of natural proof transformations:

THEOREM 4.13. *If Φ and Ψ are annotated $\boldsymbol{LK}_H^{\alpha\varepsilon}$ derivations of the same Herbrand net F, then there is a sequence $\Phi_0 = \Phi, \Phi_1, \ldots, \Phi_n = \Psi$ of derivations of F such that Φ_n differs from Φ_{n+1} by either*

—a permutation of two consecutive, non-interfering sequent rules:

—the re-association of two consecutive contraction rules

$$\dfrac{\dfrac{F, s : \exists x.A, t : \exists x.A, u : \exists x.A}{F, s + t : \exists x.A, u : \exists x.A}\, C}{F, s + t + u : \exists x.A}\, C \quad \longrightarrow \quad \dfrac{\dfrac{F, s : \exists x.A, t : \exists x.A, u : \exists x.A}{F, s + u : \exists x.A, t : \exists x.A}\, C}{F, s + t + u : \exists x.A}\, C$$

and similarly for contractions on QFFs

—the absorption of a contraction on a QFF into a tautology rule, or its reverse

$$\dfrac{\dfrac{}{G, \{i\} : P, \{i\} : P}\, i}{G, \{1\} : P}\, C \quad \longleftrightarrow \quad \dfrac{}{G, \{i\} : P}\, i$$

PROOF. Suppose Φ and Ψ are not identical sequent derivations. Then there is a branch D of Φ on which Ψ does not agree. Let ρ_0 be the last rule instance on D, counting from the root of Φ, for which Φ and Ψ agree, and let ρ', the first rule on D on which Φ and Ψ disagree, introduce the term $t : A$. Assume first (since this case is easier) that ρ' is not a contraction. Since Φ and Ψ agree up to ρ, there is a rule instance ρ_n above ρ in Ψ introducing t, with rule instances $\rho_1 \ldots \rho_{n-1}$ between ρ_n and ρ. We prove the lemma by induction on the largest such n, for any branch of Φ. First, suppose that ρ_n is a universal inference; then it can clearly be moved below ρ_{n-1}. Now suppose ρ_n is a cut. If ρ_{n-1} is a cut or an existential inference, then ρ_n can be moved below ρ_{n-1}. If ρ_{n-1} is a universal inference, then it can be moved above ρ_n if and only if its eigenvariable a is not free in the main formulae of ρ_n. But the corresponding rule to $\rho n - 1$ in Φ appears above ρ'; by strictness a cannot appear free in the premise of ρ', and so also cannot appear free in the premise of ρ_n. A similar argument works where ρ_n is an existential inference.

Now suppose that ρ' is a contraction on an existentially quantified formula, introducing an n-ary expansion $t = (w_1 + \cdots + w_n)$. We can permute the contraction inferences in Φ involving the w_i's down until they all occur, in a block, ending with ρ' – call this proof Φ'. We can do the same with Ψ, and then apply re-association and of contractions so that the contraction inferences above t is the same as in Φ' – call this proof Ψ'. Φ' and Ψ' now agree on a the block of contractions, and we may apply the induction hypothesis to find a sequence of permutations and re-associations from Φ' to Ψ'.

Finally, suppose that ρ' is a contraction on a QFF. Let S, the term ρ introduces, be a set containing indices $i_1, \ldots i_n$. As above, permute all the contractions on ancestors of S down, so they occur in a block above rho_0, both in Φ and in Ψ; call these proofs Φ' and Ψ'. The Herbrand net derived before the block of contractions

is, in both proofs: a context G and then a number of copies of each $\{i_j\}$; however, the number of copies of $\{i_j\}$ may be different in the different proofs. Now re-associate the contractions appearing in Φ' and Ψ', so that at first we only perform contractions of the form

$$\frac{G, \{i\} : P, \{i\} : P}{G, \{i\} : P}\,\mathrm{C} \tag{3}$$

Call these proofs Φ'' and Ψ''. This leads, in both proofs, to a block of contractions of the kind shown in (3), with conclusion $G, \{i_1\} : P, \ldots \{i_n\} : P$, containing only one copy of P for each tautology index. The contractions of the form shown in (3) can be pushed towards the tautology links, where they can be removed by absorbing them into the tautology. This then leaves $n-1$ instances of contraction above ρ_0, which can be re-associated so they give the same contraction tree in both proofs. □

4.4 Subnets of Herbrand Nets

We now define an analogue of the notion of subproof for Herbrand nets. While the definition of subnet is rather easy for $\mathbf{MLL}^-$ proof nets, the presence of contraction leads to a less intuitive notion for Herbrand nets.

Definition 4.14 *Subnet.* Let F be an $\alpha\varepsilon$-forest which is ACC-correct. A *subnet* of F is a subforest G of F closed under dependency (if $s \in G$ and $s \lhd t$ then $t \in G$) which itself satisfies ACC. Each root of G inherits a type from the typing derivation of the term of which it is a subterm; the *type* of a subnet is the multiset consisting of the types of its non-cut roots.

Notice that we do not require that a subnet of an Herbrand net is an Herbrand net; it might contain naked witnesses, and its indices need not yield tautologies. For example, Fig. 4 shows three subnets of the drinker's term, none of which are Herbrand nets. As another example, consider the following immediate consequence of the definition of subnet

Proposition 4.15. *Let F be an ACC-correct $\alpha\varepsilon$-forest, and let $\{i\}$ be a leaf of F. Then the subforest consisting of just the node $\{i\}$ is a subnet of F.*

There is a strong connection between subnets of an Herbrand net and subproofs of its sequentializations, which we will see once we have proved sequentialization.
The largest and smallest subnets containing a particular subterm are of particular interest:

Definition 4.16. Let F be an ACC-correct $\alpha\varepsilon$-forest, and let t be a node in F. The *empire* $e(t)$ of t in F is the largest subnet of F having t as a root. The kingdom $k(t)$ of t in F is the smallest subnet having t as a root.

The following is proved in Appendix A:

Corollary 4.17. *Every node in F has a kingdom and an empire.*

The kingdom of a node has a particular structure:

Proposition 4.18. *Let t be a node of an ACC-correct $\alpha\varepsilon$-forest F, and let G, t be its kingdom. Then the roots of G are either witnesses or cuts.*

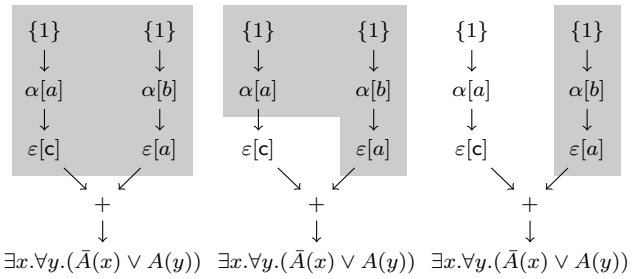


Fig. 4. Three subnets of the drinker's term

PROOF. By Prop. 4.8, if a root of G has any other form, we can find an ACC-correct subforest of G, t with t as a root, contradicting minimality of the kingdom. □

The following relation will be the key to our sequentialization and cut-elimination results.

Definition 4.19. Let F be an ACC-correct $\alpha\varepsilon$-forest. We define a relation $\ll$ on the nodes of F as follows: $t \ll s$ if $t \in k(s)$.

If t is a node of an Herbrand net F, we can think of the nodes s such that $s \ll t$ as the inference steps that *must* occur in any sequent derivation of F above the rule introducing t.

PROPOSITION 4.20. *The relation $\ll$ is a partial order on the subterms of an ACC-correct $\alpha\varepsilon$-forest.*

PROOF. See Appendix A. □

4.5 Sequentialization

We now establish that every Herbrand net arises as the conclusion of an $\mathbf{LK}^{\alpha\varepsilon}_H$ derivation. The proof that this is the case will be an induction using the following measures:

Definition 4.21. Let F be an Herbrand net.

(a) The *size* $s(F)$ of F is the number of α, ε and $\bowtie$ nodes in F.
(b) The *width* $w(t)$ of an expansion node $t = (w_1 + \cdots + w_n)$ in F is n. The *width* $w(s)$ of a propositional node $s = S$ in F is the cardinality of S.

The w-rank $w(F)$ of an Herbrand net F is $\sum_t (w(t) - 1)$, where t ranges over all expansion nodes and propositional nodes of F.

We show that all nets may be sequentialized by induction on $s(F) + w(F)$. Our base case is where $s(F) = 0$ (in which case $w(F)$ is also 0):

PROPOSITION 4.22. *If F is an Herbrand net of size* 0 *(i.e. it contains no α, ε or $\bowtie$ nodes) it is the conclusion of the tautology rule of $\mathbf{LK}^{\alpha\varepsilon}_H$.*

PROOF. Since F contains no $\bowtie$ nodes, and is a net, it can contain only one tautology index i. So F has the form $\{1\} : P_1, \ldots, \{1\} : P_n$, with $\bigvee P_i$ a tautology (since F is an Herbrand net). $\square$

In cases of non-zero measure, we look for a rule of $\mathbf{LK}_H^{\alpha\varepsilon}$ whose conclusion is F and whose premisses are also Herbrand nets – the form of the rules of $\mathbf{LK}_H^{\alpha\varepsilon}$ guarantees that the measure of each of the premisses is lower than the measure of the conclusion.

Definition 4.23. Let F be an Herbrand net, and let $t : A$ be a root of F. The root t is a *gate* of F if and only if there is a rule instance of $\mathbf{LK}_H^{\alpha\varepsilon}$, with F as conclusion, with $t : A$ as the active root in the conclusion, and with premisses that are also Herbrand nets.

If the sequent F contains a formula introduced by a universal inference rule or a contraction, then that formula is always a gate of F.

PROPOSITION 4.24. *Let F be an Herbrand net.*

(a) If $F = F', \alpha[a].t : \forall x.A$ is an Herbrand net, then $G = F', t : A[x := a]$ is also an Herbrand net.

(b) If $F = F', s_1 + s_2 : \exists x.A$, then $G = F', s_1 : \exists x.A, s_2 : \exists x.A$ is also an Herbrand net.

(c) If $F = F', S_1 \cup S_2 : P$ then $G = F', S_1 : P, S_2 : P$ is also an Herbrand net.

PROOF. Follows immediately from Prop. 4.8. $\square$

The difficulty lies in knowing when to apply the non-invertible rules of $\mathbf{LK}_H^{\alpha\varepsilon}$: the existential rule and the cut-rule. The main work of the rest of this section will be to show that each Herbrand net has a gate. We will use the notions of kingdom, empire, and the relation $\ll$, defined in the previous section. The backbone of the proof is the following characterization of the gates of an Herbrand net:

PROPOSITION 4.25. *Let $F, t : T$ be an Herbrand net*

(a) If t is of the form $\alpha[a].t$, $\{w_1, \ldots w_n\}$ or a non-singleton set S, it is a gate.

(b) if t is of the form $s_1 \bowtie s_2$ is a gate if and only if it is $\ll$-maximal.

(c) if t is of the form $(\varepsilon[M].s) : \exists x.A$, it is a gate if and only if $\varepsilon[M].s : \langle \exists x.A \rangle$ is $\ll$-maximal in $F, \varepsilon[M].s : \langle \exists x.A \rangle$.

We can immediately see that (a) holds, by Prop. 4.24. Before proving parts (b) and (c), let us observe that this characterization of gates is enough to show that every net of nonzero size has a gate:

PROPOSITION 4.26. *Let F be an Herbrand net. Either F is the conclusion of the tautology rule, or it has a gate.*

PROOF. If F has size zero and width zero, F is a conclusion of the tautology rule. Now assume that F has nontrivial size/width; by Lemma A.6, $\ll$ is a partial order on the nodes of F, so F has at least one $\ll$-maximal node t: this node is also, by definition, a root of F. If t is a gate, we are done. Suppose that t is not a gate: then by Proposition 4.25 and Proposition 4.18 it is of the form $\{i\}$ or $(\varepsilon[M].t)$. Suppose

the former: since $F = G, \{i\} : P$ has nonzero size, so does G. G is ACC-correct by Proposition 4.8: thus G has a gate $t : A$. This is also a gate of F, since $t \notin k(\{i\})$.

Finally, suppose that all $\ll$-maximal nodes of F are of the form $(\varepsilon[M_i].s_i)$, for $1 \leq i \leq n$; so

$$F = G, \quad (\varepsilon[M_1].s_1) : \exists x_1.A_1, \ldots, (\varepsilon[M_n].s_n) : \exists x_n.A_n$$

The ACC-correct $\alpha\varepsilon$-forest

$$F' = G, \quad \varepsilon[M_1].s_1 : \langle \exists x_1.A_1 \rangle, \ldots, \varepsilon[M_n].s_n \langle \exists x_n.A_n \rangle$$

has an $\ll$-maximal node, and it must be $\varepsilon[M_j].s_j : \langle \exists x_j.A_j \rangle$, for some j. This node is also $\ll$-maximal in

$$G, \quad (\varepsilon[M_1].s_1) : \exists x_1.A_1, \ldots, \varepsilon[M_j].s_j : \langle \exists x_j.A_j \rangle, \ldots, (\varepsilon[M_n].s_n) : \exists x_n.A_n,$$

(where we have placed a + below all the naked witnesses but $\varepsilon[M_j].s_j$) and so $(\varepsilon[M_j].s_j) : \exists x_j.A_j$ is a gate of F. $\square$

From this, we derive the main theorem of this section:

THEOREM 4.27 SEQUENTIALIZATION. *An annotated sequent F is an Herbrand net if and only if it is the endsequent of an $\boldsymbol{LK}_H^{\alpha\varepsilon}$ derivation π. We call π a sequentialization of F.*

PROOF. One direction is given by Prop. 4.12. For the other direction, proceed by induction on $s(F)+w(F)$. If this measure is zero, F is the conclusion of the tautology rule. Otherwise, F has a gate, and there is a sequent rule which decomposes F into one or more smaller Herbrand nets, each of which can be sequentialized by the induction hypothesis. $\square$

The following cases of Prop. 4.25 remain to be proved:

LEMMA 4.28 SPLITTING $\bowtie$. *Let $F = F', t \bowtie s : A \bowtie \bar{A}$ be ACC-correct; then $t \bowtie s$ is $\ll$-maximal in F iff there is a partition $F' = F_1, F_2$ such that $F_1, t : A$ and $F_2, s : \bar{A}$ are ACC-correct. If, further, F is an Herbrand net, then $F_1, t : A$ and $F_2, s : \bar{A}$ are Herbrand nets.*

PROOF. This is a variation on the standard "splitting tensor" theorem for **MLL** proof nets: see Section A for the proof. $\square$

LEMMA 4.29. *Let $F = G, (\varepsilon[M].t) : \exists x.A$ be ACC-correct (resp. an Herbrand net). Then $F' = G, t : A[x := M]$ is also ACC-correct (resp. an Herbrand net) if and only if $\varepsilon[M].t : \langle \exists x.A \rangle$ is $\ll$-maximal in $F'' = G, \varepsilon[M].t : \langle \exists x.A \rangle$.*

PROOF. Suppose that F is ACC-correct, and that F' is also ACC-correct, and suppose for a contradiction that $\varepsilon[M].t$ is a member of $k(X)$ for some other node X of F''. But then consider K', the kingdom of X in F'. K' is also a subnet of F'', and smaller than F since it does not contain $\varepsilon[M].t$. This contradicts minimality of the kingdom.

Suppose now that $F'' = G, \varepsilon[M].t : \langle \exists x.A \rangle$ is ACC with $\ll$-maximal node $\varepsilon[M].t : \varepsilon[M].t : \langle \exists x.A \rangle$. We show that F' is ACC. Since F' is a subgraph of F'', all its switching graphs are acyclic: we must show that they are also connected. Observe that $\mathsf{free}(M) \subseteq \mathsf{free}_\alpha(F)$. For otherwise, there is a variable a with $a \in \mathsf{free}(M)$,

$a \notin \mathsf{free}_\alpha(F)$; then there is a node of F of the form $\alpha[a].s$, and $(\varepsilon[M].t) \in k(\alpha[a].s)$, contradicting the fact that $(\varepsilon[M].t)$ is a gate. Thus the node $\varepsilon[M].t$ is connected to each switching graph only by its unique successor in the forest structure of F'', and so removing it cannot disconnect any switching graph.

Finally, notice that F and F' have the same leaves, and so each tautology index in F' gives rise to a tautology. □

The following will be useful in connecting cut-reduction in Herbrand nets with cut-reduction in $\mathbf{LK}_H^{\alpha\varepsilon}$:

PROPOSITION 4.30. *Let F be an Herbrand net, and let G be a subnet of F. Then there is a sequentialization Φ of F containing a subproof which corresponds to G in the following sense: the α, ε and cut terms of F introduced in the subproof above t are precisely those which are members of G.*

PROOF. Sequentialize F, as in the proof of the sequentialization theorem, with the caveat that no node contained in G cannot be removed: they are not considered gates of F. The algorithm will fail at the point where the remaining net H to be sequentialized has no gate to remove: all gates of H must therefore be members of G, or, in the case of a gate of the form $(\varepsilon[M].s)$, it is possible that only the witness $\varepsilon[M].s$ is a member of G. Every member of G is, of course, contained in H. On the other hand, suppose that t is a ε, α or cut node in H. Then t is contained in the kingdom of some gate s of H: but then t is a member of G, since every gate of H is a member of G, or of the form $(\varepsilon[M].s)$, where $\varepsilon[M].s$ is a member of G. □

5. CUT-ELIMINATION

The cut-free completeness of $\mathbf{LK}_H$ gives an immediate, but nonconstructive, proof of cut-elimination for Herbrand nets. In this section we will show a system of reductions ("Kingdom reduction") such that any Herbrand net may be transformed into a cut-free Herbrand net using these reductions.

Cut-reduction in sequent calculus works on subproofs. By analogy, cut-reduction on Herbrand nets works on subnets. This introduces three complications to the definition of cut-reduction. First, subnets are not necessarily Herbrand nets, and so cut-reduction will need to be defined on any ACC-correct $\alpha\varepsilon$-forest. Secondly, while the operation of replacing a subtree of a sequent proof is easy to define, it is a little harder to define replacing a subnet by its reduct, and in addition we must check that this replacement preserves correctness. Thirdly, when reducing a cut, we might have several choices of subnet to duplicate. We choose to always duplicate the *kingdom* of the $\alpha[a].s$ term in such a cut: this corresponds, in $\mathbf{LK}_H$ (by Lemma 4.30 and Theorem 4.13) to always duplicating the subproof obtained by first permuting all inferences that can be below the cut.

We turn first to the question of when we may replace a subnet F of an ACC-correct $\alpha\varepsilon$-forest with another ACC-correct $\alpha\varepsilon$-forest F'. We begin by considering replacing a subterm t of an $\alpha\varepsilon$ term $s : T$ with another term t', in such a way that we preserve typing. Clearly, if t has type R in the typing derivation of $s : T$, then replacing t with any other term with type R yields a correct typing derivation. In addition, suppose that w is a subterm of s of type $\langle \exists x.A \rangle$, and that t' has type $\exists x.A$. Then, if w appears in an expansion $r = (w + w_1 + \cdots + w_n)$ (recall that an expansion

is a formal sum, and so we can without loss of generality write w as the first term in the sum), replacing w by t' amounts to replacing r by $t' + (w_1 + \cdots + w_n)$. That is, we can replace an expansion tree by any other expansion tree with the same type, and we can in addition replace a witness of type $\langle \exists x.A \rangle$ by an expansion of type $\exists x.A$.

To replace a subnet F by another subnet F' is to replace each term of F by a corresponding term of F'. The following gadget will allow us to know when we can do that while maintaining correctness:

Definition 5.1. Let F be an ACC-correct $\alpha\varepsilon$-forest. A *substitution triple* for F is a triple $(F', f_{\text{root}}, f_{\text{taut}})$, where F' is an ACC-correct $\alpha\varepsilon$-forest, f_{taut} is a function from the tautology indices of F' to the tautology indices of F such that

$$F'_i \leftrightarrow F_{f_{\text{taut}}(i)}.$$

and f_{root} is a bijection from the non-cut roots of F to the non-cut roots of F' such that either $f(t)$ and t have the same type, or $f(t)$ has type $\langle \exists x.A \rangle$ and $f(t)$ has type $\exists x.A$.

Notice that, if F is an Herbrand net, and $(F', f_{\text{taut}}, f_{\text{root}})$ is a substitution triple for F, then F' is an Herbrand net. On the other hand, if an $\alpha\varepsilon$ forest F occurs as a subnet of an $\alpha\varepsilon$ forest G, the type-preserving properties of f_{root} allow that we may replace each root t of F by $f(t)$ in G (provided that the α bound variables of F' do not occur in G: we can guarantee this by alpha-conversion). In the following lemma, recall that Ind_F denotes the tautology indices occurring in F:

LEMMA 5.2. *Let G be an ACC $\alpha\varepsilon$-forest, and let F be a subnet of G. Let $(F', f_{\text{root}}, f_{\text{taut}})$ be a substitution triple for F. Let*

$$g_{\text{taut}} : (\text{Ind}_G \setminus \text{Ind}_F) \cup \text{Ind}_{F'} \to \text{Ind}_G$$

be the function defined as follows: $g_{\text{taut}}(i) = f_{\text{taut}}(i)$ if $i \in \text{Ind}_{F'}$, and $g_{\text{taut}}(i) = i$ otherwise. Let $G[F'/F]$ be the $\alpha\varepsilon$-forest defined as follows

—Replace each root of F with its image under f_{root};

—Replace each leaf S of G not in F with its inverse image under g_{taut}.

Let g_{root} be the obvious function from non-cut roots of $G[F'/F]$ to non-cut roots of G. Then $(G[F'/F], g_{\text{root}}, g_{\text{taut}})$ is a substitution triple for G.

PROOF. The only difficult detail to check is that $G[F'/F]$ is ACC-correct. Suppose that it is not: then there is a switching σ for $G[F'/F]$ such that the resulting switching graph is either disconnected or has a cycle. Suppose that some switching graph of $G[F'/F]$ is disconnected: then since F' is ACC correct there must be two nodes outside of F' which lie in separate components of the switching graph, from which it follows easily that some switching graph of G is disconnected. Suppose now that some switching graph of $G[F'/F]$ has a cycle. Then that cycle cannot be contained in the subnet F' of $G[F'/F]$, since F' is ACC-correct. So the cycle passes through the complement of $G[F'/F]$ and F'. Let t' and s' be two nodes of the switching graph $G[F'/F]_\sigma$ such that there is a switching path between them outside of F'. Then t', s' are either roots of F' or tautology indices found in F'. Using f_{taut} and f_{root} we can find corresponding nodes t and s, and a switching

σ' for F (which chooses t and s if their predecessors are switched, and otherwise agrees with σ) such that there is a switching path from t to s in G, outside of F. But, since t and s appear in the switching graph of F, there is also a path from t to s within F, for any switching. Thus, we find a switching cycle in a switching graph of G, contradicting that G is ACC-correct. $\square$

The substitution triples we are interested in are those that arise from the cut-reduction operations of communicating a witness and duplicating a subproof, closed under reducing in a subnet and under composition: we will call these triples *reduction-triples*.

Definition 5.3 *Reduction Triples.* The *basic* reduction triples are the following, where $F_1, \alpha[a].t : \forall x.A$ and $F_2, s : \exists x.\bar{A}$ are ACC forests and

$$F = F_1, F_2,\ \alpha[a].t \bowtie s : \forall x.A \bowtie \exists x.\bar{A} :$$

(i.e., the cut displayed *splits* F)

(a) (Identity) $(F,\ \mathbf{id}_{\text{root}},\ \mathbf{id}_{\text{taut}})$ is a reduction triple for F, where $\mathbf{id}_{\text{root}}$ and $\mathbf{id}_{\text{taut}}$ are the identity functions on the non-cut roots/tautology indices of F.

(b) (Communication) if $s = \varepsilon[M].s'$, then

$$(F_1[a := M], F_2,\ t[a := M] \bowtie s' : A[x := M] \bowtie \bar{A}[x := M], \quad f_{\text{root}}, f_{\text{taut}})$$

is a reduction triple for F, where f_{root} and f_{taut} are the evident bijections between the roots/indices.

(c) (Duplication) if s is a nontrivial expansion, if we can decompose s into $s_0 + s_1$, and if $F_1 = w_1, \ldots w_n, G$, where the w_i are witnesses and G contains only cuts, then $(F', f_{\text{root}}, f_{\text{taut}})$ is a reduction triple for F, where

$$F' = (\tau_0(w_1) + \tau_1(w_1)), \ldots, (\tau_0(w_n) + \tau_1(w_n)), \tau_0(G), \tau_1(G), F_2$$
$$\tau_0(\alpha[a].t) \bowtie s_0 : \forall x.A \bowtie \exists x.\bar{A},\ \tau_1(\alpha[a].t) \bowtie s_1 : \forall x.A \bowtie \exists x.\bar{A}$$

where f_{root} is the evident bijection between non-cut roots of F and F', f_{taut} maps indices i_0, i_1 to i if i is duplicated by the reduction, and is the identity otherwise, and τ_0, τ_1 are the renaming functions of Definition 3.11, where $V = \mathsf{free}_\alpha(F_1, \alpha[a].t)$ and I is the set of tautology indices in $F_1, \alpha[a].t$

New reduction triples can be built in two ways:

(a) (composition) If $(F', f_{\text{root}}, f_{\text{taut}})$ is a reduction triple for F, and $(F'', f'_{\text{root}}, f'_{\text{taut}})$ is a reduction triple for F', then $(F'', f'_{\text{root}} \circ f_{\text{root}}, f_{\text{taut}} \circ f'_{\text{taut}})$ is a reduction triple for F.

(b) (reduction in a subnet) If G is a subnet of F, and $(G', f_{\text{root}}, f_{\text{taut}})$ is a reduction triple for K, then $(F[G'/G], g_{\text{root}}, g_{\text{taut}})$, as defined in Lemma 5.2 is a reduction triple for F.

LEMMA 5.4. *Every reduction triple is a substitution triple.*

PROOF. It is trivial that the identity reduction triple is a substitution triple, and that the composition of two substitution triples is a substitution triple. A simple application of the ACC criterion shows that Communication and Duplication yield substitution triples – notice that in a Duplication triple f_{root} maps naked witnesses

w_i to expansions $(\tau_0(w_i) + \tau_1(w_i))$. Reduction in a subnet preserves the property of being a substitution triple, by Lemma 5.2. $\square$

As an example of the above, we will look at the reduction of a structural cut (a cut against contraction) in an Herbrand net F which does *not* split its context. This corresponds to reducing a cut in the sequent calculus which is not the last rule in the proof. For this to work, we need to find a subnet G of F containing the cut to be reduced such that the cut splits G. Such a subnet always exists: we can take the *kingdom* of the cut. The following is an immediate consequence of Prop. 4.20:

PROPOSITION 5.5. *A node t in an ACC-correct $\alpha\varepsilon$-forest F is $\ll$-maximal in $k(t)$.*

Now simply recall Lemma 4.28: a cut is splitting if and only if it is $\ll$-maximal. Let X denote the cut to be reduced. Since in X splits $k(X)$, and since all the roots of $k(X)$ are either naked witnesses or cuts, by Lemma 4.18 there is a basic reduction triple from $k(X)$ to a net K'. By a subsequent application of reduction in a subnet, we can obtain a reduction triple for F embodying a one step of cut-reduction applied to F. Since this is an important operation on Herbrand nets, we will take the trouble unpack this definition:

Definition 5.6 *The duplication reduction* DUP. Let $G = F, \alpha[a].t \bowtie_X (s_1 + s_2) : A \bowtie \bar{A}$ be an Herbrand net. Let $K = k(\alpha[a].t)$, the kingdom of $\alpha[a].t$ in G. Let V be the variables bound in α binders in K, and I be the tautology nodes in K. Let the functions τ_0 and τ_1 be renaming functions as before for the sequences V and I. Then G DUP-reduces to

$$D_a(F), \quad \alpha[x_0].\tau_0(t) \bowtie s_0 : A \bowtie \bar{A}, \quad \alpha[a_1].\tau_1(t) \bowtie s_1 : A \bowtie \bar{A},$$

where D_a is a function defined pointwise on the members of F as follows:

$$\begin{aligned}
D_a(S) &= \tau_0(S) \cup \tau_1(S) \\
D_a(t \bowtie s) &= \begin{cases} D_a(t) \bowtie D_a(s) & t \bowtie s \notin K \\ \tau_0(t \bowtie s), \quad \tau_1(t \bowtie s) & t \bowtie s \in K \end{cases} \\
D_a(\alpha[a].t) &= \alpha[a].D_a(t) \\
D_a(t_1 + \cdots + t_n) &= D_a(t_1) + \cdots + D_a(t_n) \\
D_a(\varepsilon[M].t) &= \begin{cases} (\varepsilon[M].D_a(t)) & \varepsilon[M].t \notin K \\ \tau_0(\varepsilon[M].t) + \tau_1(\varepsilon[M].t) & \varepsilon[M].t \in K \end{cases}
\end{aligned}$$

5.1 The principal lemma for partial cut-elimination

In this section we state and prove the following reduction lemma:

LEMMA 5.7. *Let $F = G,\ t \bowtie s : A \bowtie \bar{A}$ be an ACC-correct $\alpha\varepsilon$-forest, where all cuts appearing in G are of rank 0. Then F has a reduction triple $(F', f_{\text{root}}, f_{\text{taut}})$ such that F' contains only cuts of rank 0.*

This is a generalization of the following, which says that we can remove a single cut of non-zero rank from a net:

COROLLARY 5.8. *Let $F = G,\ t \bowtie s : A \bowtie \bar{A}$ be an Herbrand net, and let G contain only cuts of rank 0. There is an Herbrand net F', with the same type as F, containing only cuts of rank 0.*

PROOF. As remarked before, $(F', f_{\text{root}}, f_{\text{taut}})$ is a substitution triple for an Herbrand net F only if F' is an Herbrand net of the same type as F. □

The proof of the reduction lemma is strikingly close to Gentzen's original demonstration of cut-elimination for the classical sequent calculus, with two adjustments. These adjustments both arise from the lack of tree structure in a proof. First, we can no longer speak of the "topmost" cut in a proof; instead, we eliminate cuts which are potentially topmost:

Definition 5.9. Let F be an $\alpha\varepsilon$-forest. A cut X is an $\ll$-topmost cut of rank n in F if each cut Y with $Y \ll X$ has rank $< n$: in other words, each cut in the kingdom of X has smaller rank than X.

Second, we cannot use any notion of height as an induction measure: instead we use a more natural measure of the complexity of a cut: the number of witnesses taking place in it (its "width"). On the other hand, the proof improves on Gentzen's in that there is no need to extend the language of proofs with a *multicut* rule.

PROOF. (Of Lemma 5.1) Our proof proceeds by an induction over three measures, ordered lexicographically: the first is the size of the ACC-correct $\alpha\varepsilon$-forest, meaning the number of nodes it has. The second is the rank of the unique non-zero rank cut X appearing in the ACC-correct $\alpha\varepsilon$-forest. The final measure is the "width" of the cut: if the cut-term decorating the cut is $\alpha[a].t \bowtie s$, then the width of the cut is the width of s – otherwise the width of the cut is 0.

Our base case is where all cuts are of rank 0; there is no work to be done, and we can set $F = F'$ and both functions f_{root} and f_{taut} to be the identity.

Suppose now that X has rank n, but that F is not the kingdom of X. Then we can find a smaller ACC-correct $\alpha\varepsilon$-forest $k(X)$ containing the cut. By the induction hypothesis, we obtain a reduction triple $(K', f_{\text{root}}, f_{\text{taut}})$ for K, where K' contains only cuts of rank zero; by reduction in a subnet we obtain a reduction triple for F with the required property.

Now suppose that F *is* the kingdom of X. Then we may write F as

$$F_1, \alpha[a].t \bowtie s : \forall x.A \bowtie \exists x.\bar{A}, F_2$$

where $F_1, \alpha[a].t : A$ and $F_2, s : \bar{A}$ are also ACC, with gates $\alpha[a].t$ and s respectively. We proceed by case analysis on the structure of s.

If $s = (\varepsilon[M].s')$, there is a basic reduction triple between F and

$$E = F_1[a := M], t[a := M] \bowtie s' : A[x := M] \bowtie \bar{A}[x := M], F_2$$

which has measure less than that of F. By the induction hypothesis, there is a reduction triple $(E', g_{\text{root}}, g_{\text{taut}})$ for E, where E' contains no nonzero cuts. By composition, there is a reduction triple between F and E'.

Finally, suppose that s has the form $\varepsilon[M_1].s_1 + \cdots + \varepsilon[M_n].s_n$. Since the relation $\ll$ is a partial order on the nodes of F, there must be an $\varepsilon[M_i].s_i$ which is $\ll$-minimal among the components of s; then we can write s as $\varepsilon[M_i].s_i + s'$. There is

a basic reduction triple between F and

$$E = E',\ \alpha[a_0].t_0 \bowtie_Y \varepsilon[M_i].s_i,\ \alpha[a_1].t_1 \bowtie_Z s'.$$

Consider now the kingdom $k(Z)$ of the cut Z in E. Since we picked $\varepsilon[M_i].s_i$ to be $\ll$-minimal among the components of S, it does not appear in $k(s')$, and thus does not appear in $k(Z)$. Since $\varepsilon[M_i].s_i$ is not a member of $k(Z)$, neither is the cut Y. $k(Z)$ is, therefore, an ACC-correct $\alpha\varepsilon$-forest of lower measure than F (it contains a single cut of nonzero rank, with the same rank but lower width than the cut appearing in F) and thus by the induction hypothesis there is a reduction triple $(K', g_{\mathrm{root}}, g_{\mathrm{taut}})$ for $k(Z)$, such that K' contains only cuts of rank zero. By reduction-in-a-subnet, there is an ACC-correct $\alpha\varepsilon$-forest $E[K'/k(Z)]$ and functions h_{root} and h_{taut} forming a reduction-triple for E. The ACC-correct $\alpha\varepsilon$-forest $E[K'/k(Z)]$ now contains a single nonzero-rank cut of width 1: since $\varepsilon[M_i].s_i$ was not in $k(Z)$, the width of this cut in $E[K'/k(Z)]$ is the same as that in E. $E[K'/k(Z)]$ is thus subject to the induction hypothesis, which yields a triple $(F', h_{\mathrm{root}}, h_{\mathrm{taut}})$ for $E[K'/k(Z)]$, where F' contains pnly cuts of rank 0. We may now compose these three reduction triples to obtain the required reduction triple for F. □

As a corollary to the principal lemma, we obtain partial cut-elimination.

THEOREM 5.10 PARTIAL CUT-ELIMINATION. *Let F be an Herbrand net. There is an Herbrand net F', containing only cuts of rank zero, with the same type as F.*

PROOF. By induction on the number of nonzero-rank cuts in an Herbrand net F. If there are none, we are done. Now suppose we may remove the nonzero-rank cuts from an ACC-correct $\alpha\varepsilon$-forest containing $n-1$ nonzero-rank cuts, and let F contain n nonzero-rank cuts. Let X be a $\ll$-topmost nonzero-rank cut in F, and consider $k(X)$, it's kingdom. By the previous lemma, there is a reduction triple $(K', f_{\mathrm{root}}, f_{\mathrm{taut}})$ for $k(X)$, such that $k(X)$ contains only cuts of rank zero. The ACC-correct $\alpha\varepsilon$-forest $F[K'/k(X)]$ has the same type as F (since F has no naked witnesses), but has $n-1$ nonzero-rank cuts. Furthermore, by the properties of substitution triples every tautology index of $F[K'/k(X)]$ yields a tautology. Thus $F[K'/k(X)]$ is an Herbrand net, and we may apply the induction hypothesis to obtain an Herbrand net containing only cuts of rank zero. □

5.2 From Partial to Full cut-elimination

Usually, when one performs partial cut-elimination, it is because the remaining cuts cannot be eliminated. Here this is not the case: the cuts of rank zero may very easily be eliminated, but in a way that interferes with the notion of reduction triple. The reader might suspect that here we find a source of nondeterminism in the reductions: a term $S : P$ where S has cardinality $n > 1$, represents an $n-1$-fold contraction. Since we may form cuts $S \bowtie T$, one might expect to have to make duplications to reduce these cuts, and to have to choose a direction in which the cut should be reduced. In fact, for weak normalization we can avoid such issues, owing to the following lemma:

LEMMA 5.11. *Let $F = G,\ S \bowtie T : P \bowtie \bar{P}$ be an Herbrand net, with G cut-free: then S and T are disjoint singleton sets.*

PROOF. A simple application of the correctness that criterion: alternatively, observe that as F is an Herbrand net it must be the conclusion of an $\mathbf{LK}_H^{\alpha\varepsilon}$ derivation containing one cut, and thus two branches, each containing precisely one tautology rule. □

Such cuts are easy to eliminate

LEMMA 5.12. *Let $F, \{i\} \bowtie \{j\}$ be an Herbrand net. Then $F[i \leftarrow j]$ is an Herbrand net.*

PROOF. By induction on the height of a derivation of $F, \{i\} \bowtie \{j\}$ in $\mathbf{LK}_H^{\alpha\varepsilon}$. Since the derivation contains a cut, it cannot have height 1 - the minimal height is 2, with the proof having the form

$$\dfrac{\dfrac{}{\{i\} : P_1, \ldots, \{i\} : P_n, \{i\} : P}\, i \qquad \dfrac{}{\{j\} : Q_1, \ldots, \{j\} : Q_m, \{j\} : \bar{P}}\, j}{\{i\} : P_1, \ldots, \{i\} : P_n, \{j\} : Q_1, \ldots, \{j\} : Q_m, \{i\} \bowtie \{j\} : P \bowtie \bar{P}}\, \textsc{Cut}$$

It follows that $\bigvee_k P_k \vee \bigvee_l Q_l$ is a tautology, and so

$$\{i\} : P_1, \ldots, \{i\} : P_n, \{i\} : Q_1, \ldots, \{i\} : Q_m$$

is the conclusion of a tautology rule. The remainder of the proof is a simple induction on the height of a proof, relying on the fact that any other rule in $\mathbf{LK}_H^{\alpha\varepsilon}$ can be pushed below a cut of the form $\{i\} \bowtie \{j\}$. □

COROLLARY 5.13. *Let F be an Herbrand net containing only cuts of rank 0. Then there is an Herbrand net F' of the same type which is cut-free, which can be obtained by applying the transformation*

$$\textsc{Prop} : F, \{i\} \bowtie \{j\} \rightsquigarrow F[i := j]$$

PROOF. By induction on the number of cuts in F. Suppose that we may remove $n-1$ cuts of zero rank from a net. Then if F contains n cuts, it in particular contains one cut of the form $\{i\} \bowtie \{j\}$, which may be removed by the above lemma. The remaining proof contains $n-1$ cuts and so falls under the induction hypothesis. □

This is enough to obtain full cut-elimination for Herbrand nets. To write this theorem in a form which does not mention reduction triples, we use the defined DUP reduction from Definition 5.6: this precisely captures the kind of duplications occurring in the proof of Lemma 5.7. We will call the system of reductions comprising DUP, COMM and PROP *Kingdom reduction*, since at each stage requiring a duplication only the kingdom (the smallest possible subproof) is duplicated.

THEOREM 5.14 WEAK NORMALIZATION. *Let F be an Herbrand net with type Γ. By applying rules from Fig. 5 we may produce a cut-free Herbrand net F', also with type Γ.*

6. KINGDOM REDUCTION IS NOT CONFLUENT

Gentzen style cut-reduction is very badly behaved on proofs in classical logic. In particular, cut-reduction is highly non-confluent: the Weakening–Weakening example, due to Lafont [Girard et al. 1989] constructs, given arbitrary proofs Φ and Ψ of a sequent Γ, a third proof $\Phi * \Psi$ of Γ which reduces to both Φ and Ψ.

$$\textsc{Prop} : F, \{i\} \bowtie \{j\} \rightsquigarrow F[i := j]$$

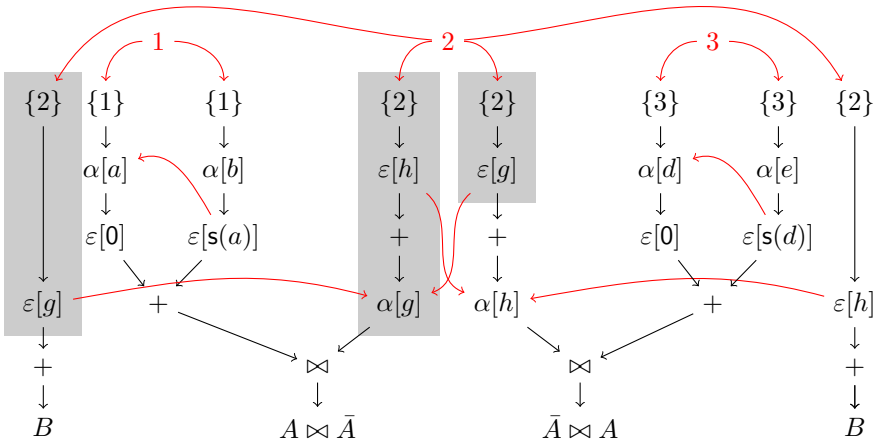
$$\textsc{Comm} : F, \quad \alpha[a].t \bowtie \{\varepsilon[M].s\} \rightsquigarrow F[a := M], \quad t[a := M] \bowtie s$$

$$\textsc{Dup} : F, \alpha[x].t \bowtie (s_0 + s_1) \rightsquigarrow D_x(F), \quad \alpha[x_0].\tau_0(t) \bowtie s_0, \quad \alpha[x_1].\tau_1(t) \bowtie s_1$$

Fig. 5. Kingdom reduction on Herbrand nets

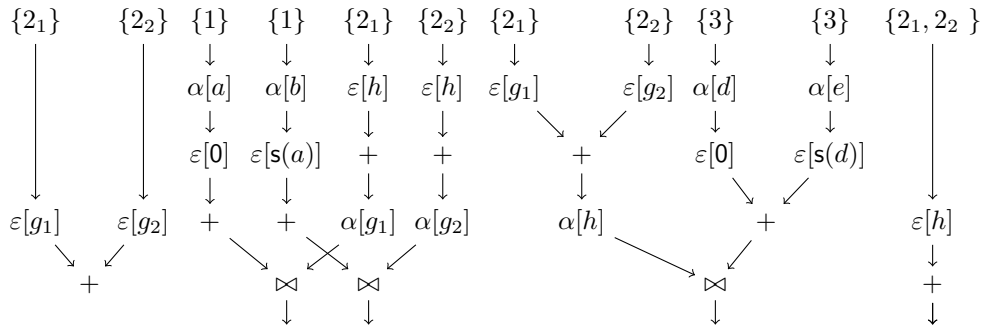
Such an easy counterexample to confluence is hard to reconstruct in Herbrand nets, as we have no weakening. We cannot even replicate the similar Contraction–Contraction example of Girard [Girard 1991], since at most one cut formula in a given nontrivial cut can be the conclusion of a contraction. Our cut-reduction system contains no critical pairs arising from the direction in which a single cut is reduced. Nevertheless, the minimal reduction system on Herbrand nets is non-confluent: the non-confluence arises between, not within, cuts: that is, the choice we are asked to make is not how to reduce one particular cut, but instead which cut we should reduce. This section is devoted to an example of this behaviour.

We work over a signature and theory axiomatizing a successor function: $\Sigma = (\mathcal{X}, \{\mathsf{0}, \mathsf{s}\}, \{\mathsf{iszero}\})$ with $\mathsf{0}$ a constant, s a unary function symbol, and iszero a unary relation symbol. The universal axiom set $\mathcal{T}$ for this theory consists of the single open formula $\neg\mathsf{iszero}(\mathsf{s}(x))$. Let A be the formula $\exists x.\forall y.(\mathsf{iszero}(x) \Rightarrow \mathsf{iszero}(y))$, and let B be the formula $\exists z.(\neg\mathsf{iszero}(\mathsf{s}(z)))$. We give a proof with cuts of the sequent B, B, containing two cuts on the formula A: depending on the order we reduce the cuts, we can obtain different witnesses above the two copies of B. Our example Herbrand net is the following:



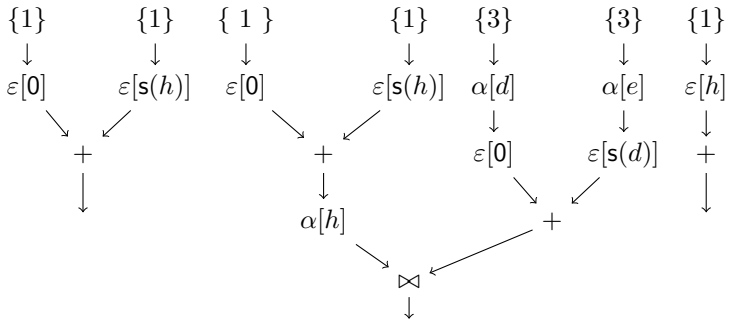
(The grey regions indicate the kingdom of the node $\alpha[g]$: we will later use this subnet to begin the elimination of cuts from this net). We leave it as a simple exercise to check that this is an Herbrand net over $\Sigma, \mathcal{T}$. To begin, we reduce the

net by a DUP-reduction applied to the left-hand cut, which duplicates the shaded subnet, the kingdom of the node $\alpha[g]$. The following net is the result:



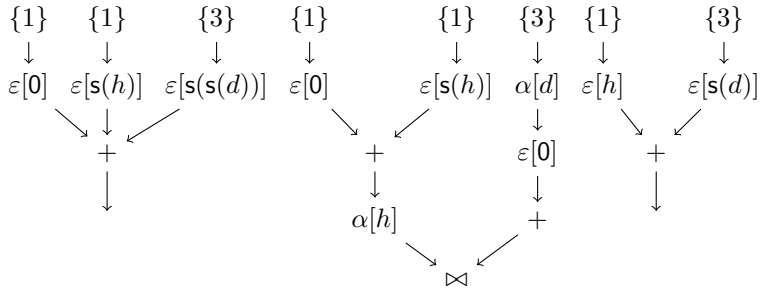
Notice that the rightmost leaf of the forest in the reduct, labelled $\{2\}$, is not in the kingdom of the cut reduced, but that the tautology index 2 is duplicated by the reduction: hence, in the reduct, this index is replaced by $\{2_1, 2_2\}$.

To continue the reduction of this net, we perform four COMM reductions, in which the ε nodes transmit their first-order terms to the corresponding α nodes. Two subsequent applications of the PROP reduction leave a net with only one cut remaining, replacing the three tautologies 1, 2_1 and 2_2 with a single tautology 1.

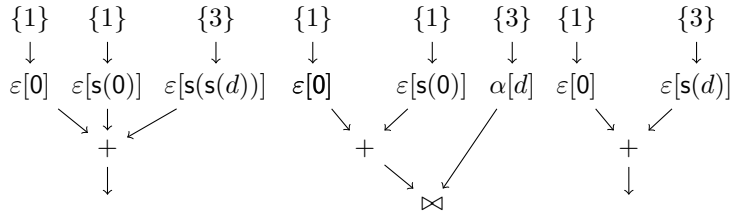


To reduce the remaining cut, we must first duplicate the kingdom of $\alpha[h]$, yielding

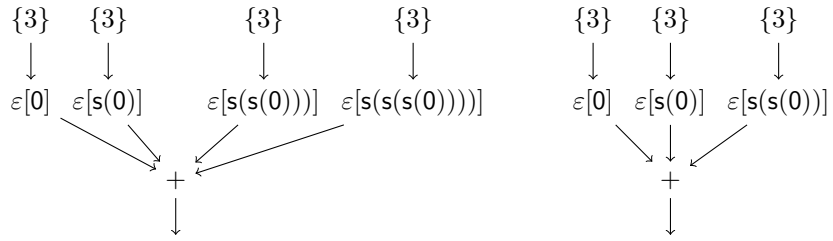
two cuts. Eliminating one of those cuts, we arrive at the following net:



We now communicate the term $\mathsf{0}$ into the eigenvariable h_1:



One application of Dup, two applications of Comm and two applications of Prop result in a cut-free net: intuitively, we substitute both of the terms $\mathsf{0}$ and $\mathsf{s0}$ for d:



We obtain a cut-free proof in which the left-hand conclusion has four witnesses, and the right-hand conclusion three witnesses. Clearly, by swapping the order in which the cuts are reduced, we could arrive at a sequence of reductions in which the left-hand conclusion has three witnesses and the right-hand four witnesses. Thus Kingdom reduction on Herbrand nets is not confluent.

6.1 The counterexample in sequent calculus

A natural question to ask is whether the phenomenon displayed by the example in the previous section relies on some property of Herbrand nets, or whether it can also be exhibited in the sequent calculus. The answer depends, of course, on what one means by cut-elimination in the sequent calculus. Proposition 4.30 tells

us that every kingdom-duplication step on a net F can be simulated in the sequent calculus: there is some sequentialization of F such that the relevant kingdom arises as a subproof. Theorem 4.13 tells us that, given enough permutations, we can freely move between those sequentializations, and thus carry out the cut-elimination steps with the sequent calculus. The counterexample given above relies on ambiguity in the order of the two cuts; in sequent calculus we are forced to choose one cut to be above the other, while in proof nets both cuts can be "topmost", in the sense that neither is contained in the others kingdom. Using the permutations induced by proof-nets one can always move the cuts past one another, but one does not *need* the full set of rule permutations to prove cut-elimination: in particular it is possible to eliminate all cuts from any $\mathbf{LK}_H$ derivation without ever permuting a cut past another cut (by always reducing a cut which is uppermost in the sequent tree). Whether or not this counterexample can be recreated in sequent calculus depends, therefore, on which proof-transformations one allows (in particular, freely moving a cut above another cut is not allowed in LK^{tq}).

7. OTHER KINDS OF REDUCTION

Kingdom duplication took some effort to define. Moreover the notion of kingdom, while natural, is little known outside the circle of specialists in proof nets. In this section we address (and reject) two seemingly natural alternatives to duplicating the kingdom, which would take less machinery to define but which are unsatisfactory for our purposes.

7.1 Copying too little: dependent subforests

Given an annotated sequent of the form

$$F, \alpha[a].t \bowtie s_1 + s_2 : A \bowtie \bar{A}$$

if we are to copy the subterm $\alpha[a].t$, to provide two copies to cut against s_1 and s_2, we must at least copy the *dependent subforest*, consisting of all the subterms t' such that $\alpha[a].t \lhd t'$ – how does that reduction behave? Since subnets are also closed under dependency, we would never copy more than the kingdom, but in general we copy much less. In addition, since the tautology jumps play no part in the dependency relation, we can simply drop them, (being sure to replace the condition on being an Herbrand net with some other tautology checking condition).

Such a reduction was studied by the author, and independently by Heijltjes (and others before us); it is seductively simple and holds the promise of an elegant abstract representation of classical proofs, but has a fatal flaw: as observed by Heijltjes [Heijltjes 2010], by duplicating dependent subforests we may reduce the example from the previous section to a forest containing a cut of the form $\alpha[a] \bowtie \varepsilon[M(a)]$, where there is a jump "across the cut". Such a "proof" can, of course, never arise as the annotation of a sequent derivation, due to strictness. This suggests, as is indeed the case, that the dependent-subforest duplicating reduction does not preserve the property of being an Herbrand net.

While we rejected this reduction in favour of Kingdom reduction, which preserves correctness with respect to the sequent calculus, Heijltjes opts in [Heijltjes 2010] instead to treat cuts with jumps across them as "garbage", and adds an extra garbage collection reduction to remove them. Since the structure at tautology

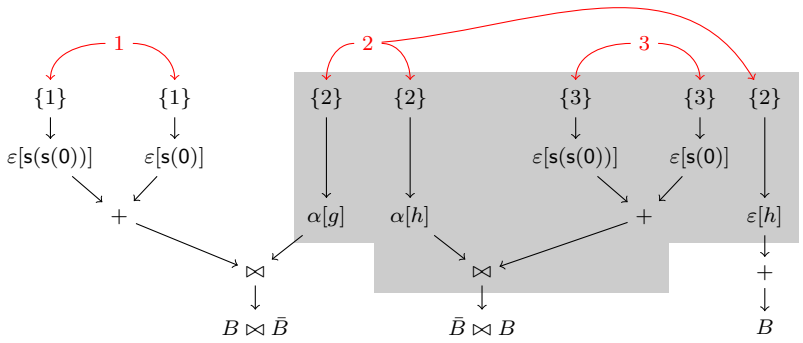


Fig. 6. A counterexample to strong normalization of Empire reduction

nodes is not needed for dependent subforest duplication, Heijltjes's "Proof Forests" can be derived from our $\alpha\varepsilon$-forests by forgetting the structure at the leaves. His correctness criterion is such that (the forgetful projection of) any Herbrand net is a correct Proof Forest. Moreover, his strategy for weak normalization seems to yield the same results as Kingdom reduction, since it always reduces an $\ll$-topmost cut (where the kingdom and dependent subforest coincide). Nonetheless, there are correct Proof Forests containing no "garbage" cuts and yet corresponding to no sequent-derivation. In the way they behave and are handled, Heijltjes's forests are rather similar to Lamarche and Strassburger's $\mathbb{N}$-nets for propositional classical logic [Lamarche and Strassburger 2005b]; in both cases, correctness with respect to sequent-calculus proofs is replaced by a weaker notion of correctness: the gain is a simpler notion of cut-reduction, but the loss is that there are "correct" proofs which do not correspond to sequential proofs.

7.2 Copying too much: empires

The very natural concept of kingdom is little-mentioned in the proof-net literature. The concept of empire, by contrast, appears in almost all introductions to the theory of proof nets for $\mathbf{MLL}^-$, and played a central role in their development. Moreover, the empire of a node is easy to calculate; for $\mathbf{MLL}^-$ nets, for example, it can be calculated in time linear in the size of the net (while calculating the kingdom is quadratic).

It is natural to ask, therefore, if this more familiar notion can be the basis of a cut-elimination for Herbrand nets. The following counterexample shows this is not possible. Let the underlying theory be as for the counterexample to confluence, and let $B = \exists z.(\neg\mathsf{iszero}(z))$. In the net shown in Figure 7.1, the shaded subnet is the *copyable part* of the empire of $\alpha[g]$; the largest subnet of the empire of $\alpha[g]$ whose roots, other than $\alpha[g]$, are all cuts or naked witnesses.

The reader can verify that, if this subnet is copied in the obvious way, and the resulting Comm/Prop redices reduced, the resulting net contains the original redex as a subnet, and indeed, it is not hard to prove that this net has no finite sequence

of reductions ending in a cut-free net, if we insist on always duplicating the empire rather than the kingdom.

8. CONCLUSIONS AND FURTHER WORK

We have shown, in this paper, a system of proof nets for classical first-order logic in prenex normal form, derived from Herbrand's theorem. The system has the minimal set of properties one might expect of a proof system for classical logic — it is sound, complete, and like Gentzen's **LK** it has weakly normalizing cut-elimination. We hope, of course, for more. Surprisingly, given the restrictions on structural rules, (and thus the avoidance of the contraction-contraction and weakening-weakening problems detailed in [Girard 1991]) cut-reduction in this system is not confluent. We seek, therefore, confluent subsystems. We conjecture, but as yet have no proof, that minimal reduction is strongly normalizing.

Similar structures to our annotated sequents arise as strategies for Coquand's game theoretical treatment of classical arithmetic [Coquand 1995]. Coquand gives a way to play a strategy containing cuts, which amounts to a non-associative composition on proofs, and it would be interesting to compare this with the non-confluent behavior of Kingdom reduction.

We look also to extend our system beyond prenex normal form, first to encompass a treatment of the propositional connectives. The papers [McKinley 2010; 2011] gives a multiplicative treatment of classical propositional proof nets which improves on [Robinson 2003] by replacing contraction (binary, defined on all formulae) by expansion (n-ary, defined only on positive formulae). It is possible to extend these nets, with the work of this paper, to full first-order logic and in addition the presentation of the axioms links can be changed so that both quantifier and axiom jumps are mediated by the α/ε of the current paper. Higher-order quantifiers could almost certainly be handled, with weak normalization being established by an adaptation of the method of reducibility candidates.

Acknowledgements The author thanks Willem Heijltjes for many stimulating and helpful exchanges, and in particular for suggesting the structure of the counterexample in Section 6. Thanks also to Michel Parigot, Lutz Strassburger, Kai Brünnler, Roman Kuznets and Stefan Hetzl for useful discussion, and to the anonymous reviewers of a previous version for their comments.

REFERENCES

Bellin, G., Hyland, M., Robinson, E., and Urban, C. 2006. Categorical proof theory of classical propositional calculus. *Theor. Comput. Sci. 364,* 2, 146–165.

Bellin, G. and van de Wiele, J. 1995. Subnets of proof-nets in MLL-. In *Proceedings of the workshop on Advances in linear logic*. Cambridge University Press, New York, NY, USA, 249–270.

Buss, S. R. 1995. On Herbrand's theorem. *Lecture Notes in Computer Science 960*, 195–209.

Coquand, T. 1995. A semantics of evidence for classical arithmetic. *J. Symb. Logic 60,* 1, 325–337.

Curien, P.-L. and Herbelin, H. 2000. The duality of computation. In *ICFP '00: Proceedings of the fifth ACM SIGPLAN international conference on Functional programming*. ACM, New York, NY, USA, 233–243.

Danos, V., Joinet, J.-B., and Schellinx, H. 1997. A new deconstructive logic: Linear logic. *The Journal of Symbolic Logic 62,* 3, pp. 755–807.

Danos, V. and Regnier, L. 1989. The structure of multiplicatives. *Archive for Mathematical Logic 28*, 181–203.

de Naurois, P. J. and Mogbil, V. 2007. Correctness of multiplicative (and exponential) proof structures is l -complete. In *CSL*, J. Duparc and T. A. Henzinger, Eds. Lecture Notes in Computer Science, vol. 4646. Springer, 435–450.

Führmann, C. and Pym, D. 2006. Order-enriched categorical models of the classical sequent calculus. *Journal of Pure and Applied Algebra 204,* 1, 21 – 78.

Führmann, C. and Pym, D. 2007. On categorical models of classical logic and the geometry of interaction. *Mathematical. Structures in Comp. Sci. 17*, 957–1027.

Gentzen, G. 1934. Untersuchungen über das logische Schließen. *Mathematische Zeitschrift 39*, 176–210, 405–431.

Girard, J.-Y. 1991. A new constructive logic: Classical logic. *Mathematical Structures in Computer Science 1,* 3, 255–296.

Girard, J.-Y. 1996. Proof-nets: The parallel syntax for proof-theory. In *Logic and Algebra*. Marcel Dekker, 97–124.

Girard, J.-Y., Lafont, Y., and Taylor, P. 1989. *Proofs and Types*. Cambridge University Press.

Heijltjes, W. 2010. Classical proof forestry. *Annals of Pure and Applied Logic 161,* 11, 1346 – 1366. Special Issue: Classical Logic and Computation (2008).

Herbrand, J. 1930. Recherches sur la theorie de la demonstration. Ph.D. thesis, Université de Paris.

Hetzl, S., Leitsch, A., Weller, D., and Woltzenlogel Paleo, B. 2008. Herbrand sequent extraction. In *Intelligent Computer Mathematics*, S. Autexier, J. Campbell, J. Rubio, V. Sorge, M. Suzuki, and F. Wiedijk, Eds. Lecture Notes in Computer Science, vol. 5144. Springer Berlin / Heidelberg, 462–477.

Hughes, D. J. D. 2006. Towards Hilbert's 24th problem: Combinatorial proof invariants. *Electron. Notes Theor. Comput. Sci. 165*, 37–63.

Lamarche, F. and Strassburger, L. 2005a. Constructing free boolean categories. In *LICS '05: Proceedings of the 20th Annual IEEE Symposium on Logic in Computer Science*. IEEE Computer Society, Washington, DC, USA, 209—218.

Lamarche, F. and Strassburger, L. 2005b. Naming proofs in classical logic. In *Proceedings of TLCA '05*. Springer-Verlag.

Lambek, J. and Scott, P. J. 1986. *Higher Order Categorical Logic*. Cambridge University Press.

McKinley, R. 2010. Expansion nets: Proof-nets for propositional classical logic. In *Logic for Programming, Artificial Intelligence, and Reasoning*, C. Fermller and A. Voronkov, Eds. Lecture Notes in Computer Science, vol. 6397. Springer Berlin / Heidelberg, 535–549.

McKinley, R. 2011. Canonical proof nets for propositional classical logic. Submitted.

Miller, D. 1987. A compact representation of proofs. *Studia Logica 46,* 4, 347–370.

Robinson, E. 2003. Proof nets for classical logic. *Journal of Logic and Computation 13,* 5, 777–797.

Strassburger, L. 2009. Some observations on the proof theory of second order propositional multiplicative linear logic. In *TLCA '09*. Springer-Verlag, Berlin, Heidelberg, 309–324.

Thiele, R. 2001. Hilbert's twenty-fourth problem. *American Mathematical Monthly 110*, 2003.

Troelstra, A. S. and Schwichtenberg, H. 1996. *Basic proof theory*. Cambridge University Press, New York, NY, USA.

A. SUBNETS OF HERBRAND NETS

The proofs contained in the appendix are very minor variations on the proofs of similar properties for $\mathbf{MLL}^-$ proof nets, as presented in [Bellin and van de Wiele 1995]. They are presented here for the sake of completeness.

The subnets of an ACC-correct $\alpha\varepsilon$-forest are closed under the following operations:

Proposition A.1. *Let G_1 and G_2 be subnets of an ACC forest.*

(a) $G_1 \cap G_2$ is a subnet of F if and only if it is nonempty.

(b) If $G_1 \cap G_2$ is nonempty, then $G_1 \cup G_2$ is a subnet.

PROOF. (a) Suppose $G = G_1 \cap G_2$ to be nonempty but not a subnet of F. It is clearly closed under dependency, so to fail to be a subnet there must be a switching σ for which G_σ is disconnected. But then either $G_{1\sigma}$ or $G_{2\sigma}$ must be disconnected.

(b) Now suppose that $G_1 \cap G_2$ is nonempty, but that $G = G_1 \cup G_2$ is not a subnet of F. Again, there must be a switching σ for which G_σ is disconnected. But since $G = G_1 \cap G_2$ is nonempty, there is a node t in G_σ present in both $G_{1\sigma}$ and $G_{2\sigma}$, and thus connected to each node of G_σ.

□

By Prop. A.1, if the set of subnets having a node t as a root is nonempty, t has an empire and a kingdom.

Definition A.2. Let F be an ACC-correct $\alpha\varepsilon$-forest, t a node of F, and σ a switching of F. Remove from F_σ the edge from t to its parent in F, if t is not a root. $F(t, \sigma)$ is the connected component of this graph containing t.

PROPOSITION A.3. *Let $e = \bigcap_\sigma F(t, \sigma)$, where σ ranges over all switchings of F and t is a node of F. Let $e(t)$ be the intersection of e with the nodes of F. $e(t)$ is a subnet of F, and t is a root of $e(t)$.*

PROOF. We must first see that $e(t)$ is closed under the dependency relation $\lhd$. This is easy to see when passing from an unswitched node to its unique successor. Suppose now that r is a switched node in $e(t)$, and that one of its immediate $\lhd$-successors s is not in $e(t)$. Then there is a switching σ such that $r \in F(\sigma, t)$ and $s \notin F(\sigma, t)$. Thus there is a path p from t to r in F_σ, and a path p' from the parent of t to s, also in F_σ. By changing the switching σ to a switching σ', where r chooses s and the parent of t chooses t (if the parent of t is switched) and leaving all other switches unchanged, we obtain a cyclic switching graph F'_σ. Hence $e(t)$ is closed under dependency.

We next observe that $e(t)$ is an ACC-correct $\alpha\varepsilon$-forest: let σ be a switching of the nodes in $e(t)$, and let σ' be an extension of that switching to F. The graph $e(t)_\sigma$ is acyclic; if not there would be a cyclic switching graph of F. To see that $e(t)_\sigma$ is connected, observe that it is the intersection of two connected graphs.

Suppose now that t is not a root of $e(t)$. Then there is a s in $e(t)$ such that $s \leq t$. Choose a switching σ_t of F such that whenever r is a switched node with $s \leq r \leq t$, we choose a switching u for r such that $u \leq t$.

Because of these choices, the unique path from t to s in F_{σ_t} uses the edge from t to its parent, and because of this does not provide a path from t to s in $F(t, \sigma_t)$. If s is in $e(t)$, then there is some other path from t to s in F_{σ_t}, but this contradicts the fact that F is correct (acyclicity of F_{σ_t}). □

PROPOSITION A.4. *The subnet $e(t)$ is the largest subnet of F having t as a root.*

PROOF. Suppose otherwise. Let G be a $\lhd$-closed subforest of F, with t as a root, which is larger than $e(t)$. Then there is a node Z of G, and a switching σ, such

that $Z \notin F(\sigma, t)$. But then there is no path from t to Z in G_σ, and so G is not ACC correct. □

The following technical lemma will be crucial:

LEMMA A.5. *Let F be an Herbrand net, and let s and t be distinct nodes of F, such that $t \in e(s)$. Let s' be the parent of s and t' the parent of t. Then*

$$s' \in e(t) \quad \textit{iff} \quad t' \notin k(s')$$

PROOF. We have that

$$G_1 = e(t) \cap k(s') \qquad G_2 = e(t) \cup k(s')$$

are ACC (since G_1 is nonempty). If $s' \in e(t), t' \in k(s')$ then G_1 has s' as a root and does not contain t', and so is a subnet with s' as a root smaller than $k(s')$ – contradiction. Similarly, if $t' \notin e(s),\ s' \notin k(t')$ then G_2 has t as a root and contains s', in contradiction of the definition of empire. □

This allows us to show that the relation $\ll$ is a partial order on the nodes of a structure.

LEMMA A.6. *Let F be an Herbrand net, and let t, s be nodes of F such that $t \ll s$ and $s \ll t$. Then $t = s$.*

PROOF. Suppose that t and s are not the same node. We have that $k(t) = k(t) \cap k(s) = k(s)$, by minimality of the kingdom.

(a) If t is an α node, or expansion node, then removing t from $k(s)$ yields a smaller subnet with s as a root, contradicting minimality of $k(s)$.

(b) If t is an ε node with unique successor t', then its kingdom is equal to $k(t') \cup \{t\}$, and so $s \in k(t')$. This contradicts the previous lemma, which says that $s \notin e(t')$. Similarly for $\bowtie$ nodes.

□