JOURNAL OF PURE AND APPLIED ALGEBRA

ELSEVIER Journal of Pure and Applied Algebra 113 (1996) 229–296

Natural deduction and coherence for weakly distributive categories

R.F. Blute[a,1], J.R.B. Cockett[b,2], R.A.G. Seely[c,*,3], T.H. Trimble[d,4]

[a] *Department of Mathematics, University of Ottawa, 585 King Edward St., Ottawa, Ont., K1N 6N5 Canada*

[b] *Department of Computer Science, The University of Calgary, 2500 University Drive N.W., Calgary, Alta., T2N 1N4 Canada*

[c] *Department of Mathematics, McGill University, 805 Sherbrooke St. W., Montréal, Qué., H3A 2K6 Canada*

[d] *Department of Mathematics, Macquarie University, Sydney, Australia*

Communicated by M. Barr; received 1 December 1992; revised 23 January 1995

Abstract

This paper examines coherence for certain monoidal categories using techniques coming from the proof theory of linear logic, in particular making heavy use of the graphical techniques of proof nets. We define a two sided notion of proof net, suitable for categories like weakly distributive categories which have the two-tensor structure (TIMES/PAR) of linear logic, but lack a NEGATION operator. Representing morphisms in weakly distributive categories as such nets, we derive a coherence theorem for such categories. As part of this process, we develop a theory of expansion–reduction systems with equalities and a term calculus for proof nets, each of which is of independent interest. In the symmetric case the expansion–reduction system on the term calculus yields a decision procedure for the equality of maps for free weakly distributive categories.

The main results of this paper are these. First we have proved coherence for the full theory of weakly distributive categories, extending similar results for monoidal categories to include the treatment of the tensor units. Second, we extend these coherence results to the full theory of $*$-autonomous categories – providing a decision procedure for the maps of free symmetric $*$-autonomous categories. Third, we derive a conservative extension result for the passage from weakly distributive categories to $*$-autonomous categories. We show strong categorical conserva-

* Corresponding author.

[1] Research partially supported by Le Fonds FCAR, Québec, and NSERC, Canada.

[2] Research partially supported by NSERC, Canada.

[3] Research partially supported by Le Fonds FCAR, Québec, and NSERC, Canada. Diagrams in this paper were produced with the help of the *TEXcad* drawing program of G. Horn and the *diagram* macros of F. Borceux.

[4] Collaboration on this work partially supported by funds from NSERC, Canada.

SSDI 0022-4049(95)00159-X

tivity, in the sense that the unit of the adjunction between weakly distributive and $*$-autonomous categories is fully faithful.

1991 Math. Subj. Class.: 03B70, 03F07, 03G30, 18D10, 19D23

0. Introduction

Weakly distributive categories were defined by Cockett and Seely in [6]. The basic structure is that of a category with two tensors, but the usual distributive law is modified to be "resource sensitive". The usual distributive law, as stated for example in [22], has an implicit asymmetry in that the number of occurrences of variables is not the same on the left side of the equation as the right. The weak distributivity corrects this. A weak distributivity is a natural transformation which acts simultaneously as a linear strength and costrength linking the two monoidal structures, and a weakly distributive category is a category equipped with two monoidal structures so linked. For more details as well as the complete definition, see [6]. This resource sensitive character is something that weakly distributive categories share with linear logic [11]. In fact, weakly distributive categories correspond precisely to multiplicative linear logic without negation. This is reflected by the fact that adding negation is precisely what is necessary to obtain $*$-autonomous categories [6, 25].

The study of weakly distributive categories is part of a program of "modularizing" linear logic: that is, allowing the buildup of the logic from as few primitives at the start as possible. Introducing negation as an initial primitive in linear logic seems rather inflexible, and certainly masks some of the underlying structure. The conservativity result of this paper shows that weakly distributive categories are the right context to study negation-free aspects of linear logic. Moreover, this program of modularization makes it possible to isolate the problem of the units which has long been a problem in the coherence theory for these categories. In the past most coherence results have had to make restrictive assumptions on the units; such restrictions are avoided with our approach.

It is natural to expect that the logical approach to coherence, originally introduced by Lambek in [19], could be exploited in this setting using linear logic. In [4, 3], Blute proposed an approach to the coherence question for various theories of monoidal categories based on a natural deduction system for linear logic, as presented graphically by proof nets [11]. The advantage of this approach (over Lambek's use of the sequent calculus) is the existence of a confluent and strongly normalizing rewrite system. This gives unique normal forms, forming the basis upon which one can approach the question of the equality of morphisms for various free monoidal categories. Furthermore, as proof nets are graphs satisfying a correctness criterion, they may be used to determine the existence of morphisms in various free monoidal categories.

To study these questions (generally referred to as "the coherence problem") we use a two-sided version of proof nets for weakly distributive categories. Since we have removed negation from multiplicative linear logic, it is no longer possible to restrict to one-sided sequents; we must have proof nets with both premises and conclusions. We have also found it convenient to use a graphical notation similar to the Joyal–Street tensor calculus [15]. A consequence of using two-sided nets is the absence of an explicit cut link; cut elimination is replaced by normalization, as in more traditional natural deduction systems such as those studied in [24].

It is straightforward to introduce the units, even without changing the correctness criterion. This straightforward approach, unfortunately, proves insufficient for coherence: it is possible for nets with the same expanded normal form to correspond to different morphisms, and for nets with different expanded normal forms to correspond to the same morphism. However, by adapting ideas developed by Trimble in [28], we give an alternate characterization of the Lambek equivalence relation in a form more convenient for our nets-based presentation. The essence of this is contained in the "Rewiring Theorem" (Theorem 3.3). Furthermore this led to the development of a calculus of rewiring (Section 3) based on a series of rewrite equations, most of which are more local in nature than the Rewiring Theorem – in essence the non-localness has been isolated in a small subset of rewirings. As an application, a full coherence result for weakly distributive categories (with units) is obtained, without the need for any of the restrictions frequently placed on the units in such coherence theorems.

Furthermore, following the extension of weakly distributive categories to $*$-autonomous categories presented in [6], we extend these results to proof nets for $*$-autonomous categories, and hence for multiplicative linear logic (with units); so we get an improvement of the coherence theorem for $*$-autonomous categories [3]. As a consequence of this, we are able to show that full multiplicative linear logic is a conservative extension of weakly distributive logic, in the strong sense that the unit of the adjunction between weakly distributive and $*$-autonomous categories is fully faithful. Conservative extension results are often used as a pretext to work exclusively in the richer setting. However, they are also a reminder of the possibility that the extra baggage of the richer setting was not necessary for the result at hand – which could perhaps be better stated in the weaker setting.

We have found it preferable to adopt a somewhat different approach to net-theoretic matters than has become standard in linear logic, and some of those differences we mention here. First, we replace the net criterion (as defined by Danos and Regnier [8] for instance) with a more local algorithm for determining the sequentiality of a proof structure. This algorithm is essentially the polynomial-time algorithm of Danos [7]. One reason for doing this is that the usual net criterion is insufficient in the non-commutative case with units – we plan to discuss the various issues dealing with the non-commutative case in a sequel. Furthermore, our proof of sequentialization applies to the case, essential to the categorical context, with added non-logical axioms – unlike Girard's original proof. Essentially the same approach has been discovered independently by Lafont [18]. Similarly, the empire criterion in Trimble's original Rewiring Theorem has been replaced by moves which are more local in nature. We have structured the graph rewriting rules we need so as not to assume commutativity of the tensors as much as possible. Most of our results can be lifted to the non-commutative case, and we have indicated where differences in the treatment arise. We have found that examining what happens to proof nets once one allows non-logical axioms and once one allows non-commutative tensors is very illuminating. We have introduced a term calculus for proof nets so as to make the rewrite rules more precise and to facilitate an implementation of the ideas of this paper. In Appendix B we include a description of expansion–reduction rewrite systems with equality which develops the theory somewhat further than is usual in the standard literature.

Some of the matter of this paper has been treated elsewhere as part of the study of linear logic. We mention only a few references here. The correctness criterion for nets with units is included as a footnote in [13], and is mentioned in [12]. In [1] essentially the same proof nets are defined for the planar unit free case. A coherence result in the case where the units coincide was obtained by Soloviev [26]. Finally, we direct the reader's attention to the paper [20] by Lambek, in which a more complete survey of the historical development of these notions is given.

1. Weakly distributive categories and categorical preliminaries

In this section, we review some categorical definitions, and briefly consider the implications for proof nets of having symmetric or non-symmetric tensors.

In [6] weakly distributive categories were introduced; for our purposes we may take the following as the definition. [5]

Definition 1.1 (*weakly distributive category*). A *weakly distributive category* **C** is a category with two tensors and two weak distribution natural transformations. The two tensors will be denoted by $\otimes$ and $\oplus$ and we shall call $\otimes$ the *tensor* and $\oplus$ the *cotensor*. Each tensor comes equipped with a unit object, an associativity natural isomorphism, and a left and right unit natural isomorphism:

$$(\otimes, \top, a_\otimes, u^L_\otimes, u^R_\otimes)$$

$$\begin{aligned} a_\otimes &: (A \otimes B) \otimes C \longrightarrow A \otimes (B \otimes C) \\ u^R_\otimes &: A \otimes \top \longrightarrow A \\ u^L_\otimes &: \top \otimes A \longrightarrow A \end{aligned}$$

$$(\oplus, \bot, a_\oplus, u^L_\oplus, u^R_\oplus)$$

$$\begin{aligned} a_\oplus &: (A \oplus B) \oplus C \longrightarrow A \oplus (B \oplus C) \\ u^R_\oplus &: A \oplus \bot \longrightarrow A \\ u^L_\oplus &: \bot \oplus A \longrightarrow A \end{aligned}$$

[5] We shall follow the usage of [6] in this paper, using $\oplus$ for the cotensor (which corresponds to Girard's "par", viz. ⅋ in Girard's notation). Note then that $\oplus$ is not the categorical coproduct.

The two weak distribution transformations shall be denoted by:

$$\delta_L : A \otimes (B \oplus C) \longrightarrow (A \otimes B) \oplus C$$
$$\delta_R : (B \oplus C) \otimes A \longrightarrow B \oplus (C \otimes A).$$

A symmetric weakly distributive category is a weakly distributive category, both of whose tensors are symmetric.

This data must satisfy standard coherence conditions, discussed in [6], which we shall not repeat here, as this paper proposes to replace them with graph rewrites.

The proof nets to be introduced in the next section are suitable for the symmetric case. However, the non-symmetric case is very natural from the net viewpoint as then the sound nets are just those that are planar; by "planar" we just mean that there are no crossings in the graph.

One of the ideas that pushed the development of the notion of weakly distributive category was that this notion helped us understand the structure of a polycategory. Polycategories were introduced (by Lambek and by Szabo [27]) to give a categorical description of the proof-theoretic structure of classical logic, and may be briefly described as "categories whose arrows have many objects as source and many objects as target". These source and target objects correspond to the premises and conclusions of a sequent in classical logic. Since linear logic can also be described in terms of such sequents, polycategories seemed to be a natural context to investigate the negation-free fragment of multiplicative linear logic. However, the notion of a polycategory is unwieldy, and rather unfamiliar. The notion of a tensor (or monoidal) category is much more familiar, and so it was important to arrive at a categorical understanding of polycategories. In [6] (where a precise definition of polycategory may be found) it was shown that weakly distributive categories are precisely the categorical analogue to polycategories, the two tensors corresponding to the interpretation of the "commas" on either side of a sequent. The weak distributivities are precisely what is necessary to define polycategorical composition; in other words, they are precisely what is needed to interpret the cut rule. More importantly from our perspective (dwelling as it does on coherence questions) the defining polycategorical equations correspond exactly to the coherence conditions for weakly distributive categories.

A further observation made in that paper was that adding negation in the most naive manner was sufficient to capture the structure of $*$-autonomous categories. For the symmetric case the following suffices.

Definition 1.2. A symmetric weakly distributive category with negation is a symmetric weakly distributive category with an object function $(\_)^\perp$, together with the following parameterized families of maps ("contradiction" and "tertium non datur"):

$$A \otimes A^\perp \xrightarrow{\gamma_A} \perp \qquad \top \xrightarrow{\tau_A} A^\perp \oplus A$$

which satisfy the following coherence condition

$$\begin{tikzcd}
\top \otimes A \arrow[r, "\tau \otimes i"] \arrow[dd, "u_\otimes^L"'] & (A \oplus A^\perp) \otimes A \arrow[d, "\delta_R^R"] \\
= & A \oplus (A^\perp \otimes A) \arrow[d, "i \oplus \gamma"] \\
A & A \oplus \perp \arrow[l, "u_\oplus^R"]
\end{tikzcd}$$

and the dual diagram for $A^\perp$.

Theorem 1.3 (Cockett and Seely [6]). *The notions of symmetric weakly distributive categories with negation and $*$-autonomous categories coincide.*

The simplicity of this extension (just a few non-logical axioms and a few new equations added) allows this to fit into the context of expansion–reduction rewriting systems with equalities, and so we shall derive coherence for $*$-autonomous categories from our result for weakly distributive categories.

2. Proof nets

In this section we introduce a two-sided notion of proof net in order to represent the proofs of the negation-free fragment of mLL (multiplicative linear logic). The proof theory of this negation-free fragment of mLL corresponds to the doctrine of weakly distributive categories, see [6], therefore these proof nets also can be used to represent the maps of a weakly distributive category.

The purpose of this section is to establish the connection between these proof nets, the two-sided sequents of the negation-free fragment of mLL and the morphisms of a weakly distributive category. We shall work with both the commutative logic (with exchange rule) and the non-commutative logic (lacking the exchange rule and with a restricted cut). These correspond on the categorical side to respectively symmetric weakly distributive categories and non-symmetric weakly distributive categories.

For those familiar with the usual notion of proof nets (see [11, 8] for example), it is worth indicating the modifications we have made to that notation. Graphically we represent logical formulae as "wires" (edges) and logical rules as "components" (nodes). This, rather suggestively, makes the nets look like circuits. Furthermore, as we are dealing with a negation-free fragment, we must modify the usual proof nets so that negated conclusions (i.e. ones of the form $C^\perp$) become unnegated premises. Logically this corresponds to using a two-sided sequent proof system.

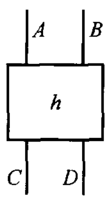
More fundamentally we have found it necessary to have an underlying syntax for the proof structures. This is because the graphical notation, while very natural and convenient, can obscure some crucial details. This syntax also serves as a term logic.

While the main purpose of the paper is to treat commutative proof theory (coherence for symmetric weakly distributive categories), the basic machinery developed works also for the non-commutative proof theory (non-symmetric weakly distributive categories) and we shall develop the non-commutative case in parallel to provide further perspective.

2.1. Circuits

Preliminary to introducing the definition of a proof net it is important to understand what is meant by a *typed circuit*. To build a typed circuit one needs a set of *types* $\mathcal{T}$ and a set of *components* $\mathcal{C}$. Each component $f \in \mathcal{C}$ has a signature $\mathrm{sig}(f) = (\alpha, \beta)$, a pair of lists of types, where α is the types of the *input ports* and β the types of the *output ports*. These give the types of the variables and covariables (a term introduced by Lambek) of the component f. We require that every component has at least one of α and β non-empty: this ensures that they can interact with the external world.

A component, therefore, is to be thought of as a black box with a number of (typed) input ports and output ports:

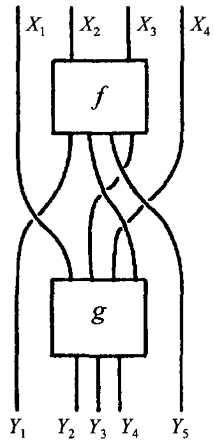


One can attach variable terms or wires to the ports of a component to obtain a *primitive circuit expression.* This we may write as $[x_1,x_2]h[y_1,y_2]$ where each wire must be of the correct type for its corresponding port (so $x_1 : A$, viz. x_1 is of type A, $x_2 : B$, $y_1 : C$, and $y_2 : D$) and the wire names in each list must be distinct. The resulting circuit expression has a list of input wires $\{x_1,x_2\}$ and a list of output wires $\{y_1,y_2\}$. Wire names are to be viewed as variables of the appropriate type.

One can plug (primitive) circuit expressions together to form new circuit expressions by juxtaposition:

$$([x_2,x_3]f[y_1,z_1,y_5,z_2]; [x_1,z_2,x_4,z_1]g[y_2,y_3,y_4]).$$

Note that although the sublist z_1,z_2 of the output of f is not the same as the sublist z_2,z_1 of the input of g (they are equal as sets, of course), we can still join these circuits at these wires. In this sense, we treat the lists of input and output wires as sets (in the symmetric case – in the non-symmetric case we shall disallow such juxtaposition, as we discuss below).



Notice how we draw the circuits with their input types listed across the top and their output types across the bottom. In general we shall not name the wires but may indicate the types of internal wires when we deem it helpful.

The wires' names common to the output of the first component and the input of the second become bound in this juxtaposition and indicate the configuration of the connection of the components. Notice that each such wire is to connect a single output and input port. To perform a legal juxtaposition $V_f f W_f; V_g g W_g$ we must avoid variable clashes: specifically the input variables of f, V_f must be disjoint from $V_g - W_f$ and similarly the output variables of g, W_g have to be disjoint from $W_f - V_g$. The result, in this case, is a new circuit expression with internal (or bound) wires $\{z_1, z_2\}$, free input wires $\{x_1, x_2, x_3, x_4\}$ and free output wires $\{y_1, y_2, y_3, y_4, y_5\}$.

If the wires in juxtaposition are always renamed to avoid any variable-name clashes this juxtaposition operation is associative (notice that in particular the bound variables may have to be renamed to avoid a clash as they can become bound at different stages in building reassociated compositions). Furthermore, in the obvious manner, one component can be exchanged with or commuted past another whenever no wires are bound in their juxtaposition.

Notice that we have allowed the wires z_1 and z_2 to "cross" and indeed to access x_1 and x_4 as inputs also requires crossings. The ability to cross wires in this manner corresponds to the commutativity of the underlying logic. To obtain the non-commutative juxtaposition, or the *planar juxtaposition*, we properly treat the wires as non-repeating lists instead of imagining them as sets; this requires that we alter the criteria for juxtaposition. A legal juxtaposition then involves thirteen non-repeating lists: the input and output wires of each component, the wires which become bound V', or through which the two components interact (which must contain all the variables in common between W_g and V_f), two sublists of W_f, W_f^L and W_f^R, two sublists of V_g, V_g^L and

V_g^R, two lists of straight through wires U_L and U_R, and $V_{f;g}$ and $W_{f;g}$ the input and output list for the juxtaposition. These must satisfy the following conditions (_@_ is the concatenation operation):

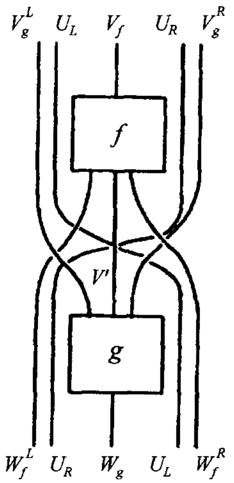
$$W_f = W_f^L @ V' @ W_f^R$$

$$V_g = V_g^L @ V' @ V_f^R$$

$$V_{f;g} = V_g^L @ U_L @ V_f @ U_R @ V_g^R$$

$$W_{f;g} = W_f^L @ U_R @ W_g @ U_L @ W_f^R$$

at most one of V_g^L and W_f^L is non-empty, at most one of V_g^R and W_f^R is non-empty. If U_L is non-empty then V', V_g^R, W_f^L, and U_R must be empty. If U_R is non-empty then V', V_g^L, W_f^R, and U_L must be empty. The reasons for these restrictions are illustrated below. The wires represent the lists of wires above: when two wires cross at most one list can be non-empty.



Notice that $V_{f;g}$ and $W_{f;g}$ are only uniquely determined by the circuits being composed when some wires become bound: that is V' is a non-empty list by virtue of some of the output wires of f becoming captured as inputs to g. When there are no such bound wires not only can the circuits be side by side (in two ways) but they can also float apart and allow wires (from other components) to run between them. The composition is uniquely determined, however, when the input list $V_{f;g}$ and output list $W_{f;g}$ are given.

To determine whether circuit expressions can be composed it suffices to perform all the interacting compositions which can be produced by reassociating and exchanging. A composition is *interacting* in case it binds some non-empty set of variables. Interacting compositions uniquely determine larger circuit expressions which can be used in further interacting compositions. Eventually, a stage will be reached when it is impossible to utilize the associativity and exchange rule to produce further interacting compositions. One can then use the fact that non-interacting components (provided their inputs and outputs are distinct) always legally compose. The precise form of their composition, of course, cannot be determined without specifying the arrangement of the external wires. Such specification introduces a further important circuit construct.

A (non-planar) circuit expression C can be *abstracted* by sandwiching it between a non-repeating list of input wires and a non-repeating list of output wires. This is written $\langle x_1,\dots,x_n|C|y_1,\dots,y_m\rangle$. Furthermore one can indicate the types of the input and output wires by the notation

$$T_1,\dots,T_n : \langle x_1,\dots,x_n|C|y_1,\dots,y_m\rangle : T'_1,\dots,T'_m$$

For an abstraction to be *closed*, all the free input wires of C must occur in the abstracting input wire list and all the free output wires of C must occur in the abstracting output list. Furthermore, any wire in the abstracting input list which is not a free input of C must occur in the abstracting output list and, similarly, any wire in the abstracting output list which is not a free output of C must occur in the abstracting input list. We shall be concerned here with closed abstractions.

In particular, we shall allow ourselves to use this technique of abstracting to isolate a wire (or many wires) as $T : \langle x|\emptyset|x\rangle : T$, where $\emptyset$ is the empty circuit and the unit for juxtaposition. This is to be regarded as the "identity map" on the type T: the ability to abstract (and the existence of an empty circuit) are important when we consider how to form categories from circuits.

When a circuit expression is abstracted in this fashion all the wire names become bound; externally an abstraction presents only the lists of types of the inputs and outputs. This permits an abstracted circuit expression to be used as if it were a primitive component. An abstraction used as a component is equivalent to the circuit obtained by removing the abstraction with a substitution of wires *outward* (with the prior renaming of the bound internal wires away from the external wires so as to avoid capture or variable clash). To see why the variables of the abstraction are used to substitute the external wires it suffices to consider the use of the "identity" abstraction mentioned above (or indeed any abstraction with "straight-through" wires). This operation of removing an abstraction we call abstraction *dissipation*; it is essentially β-reduction. The reverse operation is to *coalesce* an abstraction. These operations become particularly important when we consider how one adds identities to the basic circuit identities.

To obtain a closed *planar abstraction* we must insist on preserving the order of the wires. If V is a valid input, and W a valid output list for C then $\langle V'|C|W'\rangle$ is a valid *planar abstraction* in case there are two non-repeating lists of wires, L and R, such that $V' = L@V@R$ and $W' = L@W@R$.

We may now define the notion of a circuit based on a set of components:

Definition 2.1 (*Circuits*). (i) (Planar) *$\mathscr{C}$–circuit expressions* are generated by:
- The empty circuit, $\emptyset$, is a circuit expression.
- If c_1 and c_2 are circuit expressions which can be legally juxtapositioned (see above) then $c_1; c_2$ is a circuit expression.
- If $f \in \mathscr{C}$ is a component with $\mathrm{sig}(f) = (\alpha, \beta)$ and V is a non-repeating wire list with type α and W is a non-repeating wire list with type β then VfW is a circuit expression.
- If F is an abstracted circuit with signature $\mathrm{sig}(F) = (\alpha, \beta)$ and V is a non-repeating wire list with type α and W is a non-repeating wire list with type β then VFW is a circuit expression.

(ii) A (planar) *circuit* is a closed (planar) abstracted circuit expression.

One circuit expression (and by inference circuit) is equivalent to another precisely when one can obtain the second from the first by a series of the following operations:
- Juxtaposition reassociation (with possible bound variable renaming to avoid clashes), $c_1; (c_2; c_3) = (c_1; c_2); c_3$;
- Empty circuit elimination and introduction, $c; \emptyset = c = \emptyset; c$;
- Non-interacting subcircuits exchange, $c_1; c_2 = c_2; c_1$;
- Renaming of bound variables;
- Abstraction coalescing and dissipating.

The fact that circuit equivalence under these operations is decidable is immediately obvious when one presents them graphically. Indeed, while it is nice to have a syntax for circuits it is very much more natural and intuitive to simply draw them!

The $\mathscr{C}$-circuits, besides permitting these standard manipulations, can also admit additional manipulations. These take the form of additional identities specified as equalities $c_1 = c_2$ between (closed) abstracted circuits with the same signature. To apply such an identity to a circuit, it is necessary to be able to coalesce c_1 (up to α-conversion) within the circuit. Once this has been done one can replace c_1 with c_2. Diagrammatically this corresponds to a surgical operation of cutting out the left-hand side and replacing it with the right-hand side: accordingly such additional identities will be referred to as *rules of surgery*.

In the remainder of the paper we shall try to maintain intuition by providing the diagrammatic form of the more important circuits but will rely on the circuit notation for the detailed exposition.

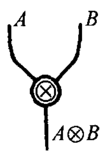
2.2. *Proof nets*

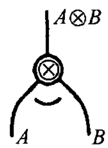
The particular typed circuits in which we are interested are $(\otimes, \oplus)$-circuits with components $\mathscr{C}$ and *atomic types $\mathscr{A}$*. These have as types the *positive linear formulae*: that is, starting with the set $\mathscr{A}$ of atomic types the inductively defined set:
- $A \in \mathscr{A}$ is a formula,
- if A and B are formulae then $A \otimes B$ and $A \oplus B$ are formulae,
- $\top$ and $\bot$ are formulae.

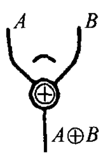
The set of components $\mathscr{C}$ is as before. However, there are in addition ten polymorphic components; we shall call these components links to distinguish them from arbitrary components:

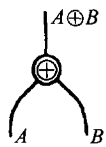
$[A,B] \otimes I[A \otimes B]$	$\otimes$-introduction
$[A \otimes B] \otimes E[A,B]$	$\otimes$-elimination
$[A,B] \oplus I[A \oplus B]$	$\oplus$-introduction
$[A \oplus B] \oplus E[A,B]$	$\oplus$-elimination
$[\,]\top I[\top]$	unit introduction
$[A,\top]\top E^R[A]$	unit right elimination (thinning)
$[\top,A]\top E^L[A]$	unit left elimination (thinning)
$[A]\bot I^R[A,\bot]$	counit right introduction (cothinning)
$[A]\bot I^L[\bot,A]$	counit left introduction (cothinning)
$[\bot]\bot E[\,]$	counit elimination

These are drawn as follows:

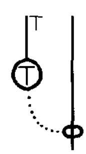
$(\otimes I)$


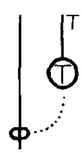
$(\otimes E)$


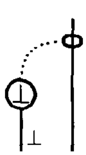
$(\oplus I)$


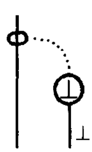
$(\oplus E)$


$(\top I)$


$(\top E)^L$


$(\top E)^R$


$(\bot I)^L$


$(\bot I)^R$


$(\bot E)$


Proof theoretically the $\otimes I$ link corresponds to the right-introduction rule for the tensor and $\otimes E$ to the left-introduction rule of the tensor.

$$(\otimes R)\ \frac{\Gamma_1 \longrightarrow \Gamma_2, A, \Gamma_3 \quad \Delta_1 \longrightarrow \Delta_2, B, \Delta_3}{\Gamma_1, \Delta_1 \longrightarrow \Gamma_2, \Delta_2, A \otimes B, \Gamma_3, \Delta_3}, \qquad (\otimes L)\ \frac{\Gamma_1, A, B, \Gamma_2 \longrightarrow \Gamma_3}{\Gamma_1, A \otimes B, \Gamma_2 \longrightarrow \Gamma_3}$$

Thinking in terms of natural deduction, these rules induct a (bijective, once we have the right equations) correspondence indicated by these "rules":

$$\frac{\Gamma_1, A \otimes B, \Gamma_2 \vdash \Delta}{\Gamma_1, A, B, \Gamma_2 \vdash \Delta}\ (\otimes\text{-intro}), \qquad \frac{\Gamma_1, A, B, \Gamma_2 \vdash \Delta}{\Gamma_1, A \otimes B, \Gamma_2 \vdash \Delta}\ (\otimes\text{-elim})$$

Dually, $(\oplus I)$ and $(\oplus E)$ correspond respectively to introduction and elimination rules for the cotensor. In fact, the idea is that one should be able to translate from a sequent proof into a circuit with the above links. A component $f \in \mathscr{C}$ will correspond to a "non-logical axiom" (the inputs will therefore be tensored together while the outputs are cotensored).

2.3. *Why thinning links?*

A curiosity is the apparent lack of symmetry between unit introduction and elimination links. Logically they correspond to the (again bijective) correspondence

$$\frac{\Gamma_1, \top, \Gamma_2 \vdash \Delta}{\Gamma_1, \Gamma_2 \vdash \Delta}\ (\top\text{-intro}), \qquad \frac{\Gamma_1, \Gamma_2 \vdash \Delta}{\Gamma_1, \top, \Gamma_2 \vdash \Delta}(\top\text{-elim})$$

which may be derived in the commutative logic with cut from the sequent rules

$$(\top L)\ \frac{\Gamma_1, \Gamma_2 \vdash \Gamma_3}{\Gamma_1, \top, \Gamma_2 \vdash \Gamma_3} \qquad (\top R) \qquad \vdash \top$$

$$(\bot L) \qquad \bot \vdash \qquad\qquad (\bot R)\ \frac{\Gamma_1 \vdash \Gamma_2, \Gamma_3}{\Gamma_1 \vdash \Gamma_2, \bot, \Gamma_3}$$

For example we might have expected the unit elimination link to be simply $[\top]\top E[\,]$. As this is a rather crucial aspect of this development some preliminary discussion of this is in order.

We shall want the following circuit identity:

$$\top : \langle z_0 | [\,] \top I[z']; [z', z_0] \top E^L[z_1] | z_1 \rangle : \top = \top : \langle z || z \rangle : \top.$$

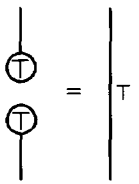
which corresponds to the equivalence of the following proof (which uses a cut on the left $\top$) to the identity:

$$\frac{\vdash \top \qquad \dfrac{\top \vdash \top}{\top, \top \vdash \top}}{\top \vdash \top}$$

If we had let the unit elimination link be as suggested above this identity would become

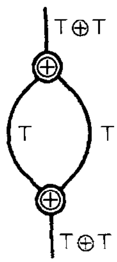
$$\top : \langle z|[z]\top E[\,];[\,]\top I[z]|z\rangle : \top = \top : \langle z||z\rangle : \top.$$

Graphically,

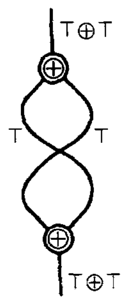


This will not do, however – the lack of a thinning link here is fatal to the coherence questions we are concerned with.

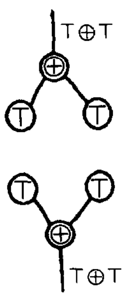
To illustrate why thinning links are important consider the following simple example which compares the identity with the cotensor twist map applied to the counit tensored with itself:



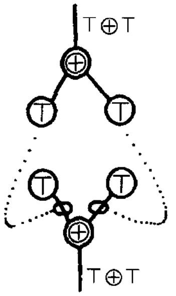
and



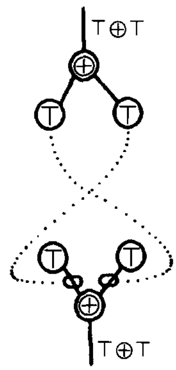
Given the above identity without thinning links these would both be equivalent to the same net:



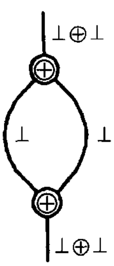
The second net would lose the twist because of the disconnection. However, the twist and the identity are not equivalent as morphisms. For example, **Sets** is a weakly distributive category with respect to the product and coproduct, but in this category clearly the twist on the two-element set is not the identity. The point is with thinning links we can at least distinguish these maps as nets, as we see below.



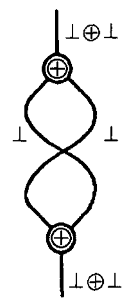
and



so there is hope that we can arrange for them to be inequivalent. Note, however, the different behavior of the units: if we replace $\top$ with $\perp$, then these nets do correspond to equivalent derivations, since $\perp \oplus \perp$ is isomorphic to $\perp$ and the identity is the same as the twist of $\perp$. Thus, these two nets when we expand the $\perp$ identity wires in the same manner must be equivalent.



and



To make these circuits equivalent it is clear that we must be able to rewire the thinning links in some manner. It is not too surprising that such an ability will be required: thinning links merely indicate a point at which a unit (or counit) has been introduced by thinning – there is considerable inessential choice going on here. For example, consider the three sequent calculus derivations of the sequent $A, \top, B \longrightarrow A \otimes B$ obtained by thinning in each of the possible places (these clearly must be equivalent – they represent the same morphism):

$$\frac{\dfrac{A \longrightarrow A \quad B \longrightarrow B}{A, B \longrightarrow A \otimes B}}{A, \top, B \longrightarrow A \otimes B} = \frac{\dfrac{A \longrightarrow A}{A, \top \longrightarrow A} \; B \longrightarrow B}{A, \top, B \longrightarrow A \otimes B} = \frac{A \longrightarrow A \; \dfrac{B \longrightarrow B}{\top, B \longrightarrow B}}{A, \top, B \longrightarrow A \otimes B}$$

As circuits, these are just the $(\otimes I)$ link with a $(\top E)$ link attached to the three possible nodes.

2.4. Sequentialization

In order for a $(\oplus, \otimes)$-circuit to be a representation of a proof it is necessary to be able to collect the circuit into a sequent. The circuits which can be rewritten into a single sequent by a series of directed surgeries are called *sequential* as the process of collecting them, sequentialization, also demonstrates that they represent a sequent calculus proof of the positive fragment of mLL.

We may arrange the process of sequentialization as an expansion/reduction rewriting system (in the sense of Appendix B) on the circuit which tries to translate the net into a sequent. There are a number of subtleties concerning this translation to which we will return, however, the basic idea is that if one can rewrite the circuit into a single sequent then one will have established its sequential nature.

We start with the translation rules for planar circuits (we abuse our notation by dropping some of the typing information where it may easily be re-inserted):

Expansion rule:

$$A : \langle x|\emptyset|x\rangle \Rightarrow \langle x|[x]\boxed{A \vdash A}[x]|x\rangle : A$$

Reduction rules:

$$: \langle|[\,]\top I[z]|z\rangle$$

$$\Rightarrow \langle|[\,]\boxed{\ \vdash \top}[z]|z\rangle : \top$$

$$\bot : \langle z|[z]\bot E[\,]|\rangle$$

$$\Rightarrow \langle z|[z]\boxed{\bot \vdash\ }[\,]|\rangle :$$

$$A, B : \langle x, y|[x, y] \otimes I[z]|z\rangle$$

$$\Rightarrow \langle x, y|[x, y]\boxed{A, B \vdash A \otimes B}[z]I[z]|z\rangle : A \otimes B$$

$$A \oplus B : \langle x|[x] \oplus E[y, z]|y, z\rangle$$

$$\Rightarrow \langle x|[x]\boxed{A \oplus B \vdash A, B}[y, z]|y, z\rangle : A, B$$

$$A, \top : \langle x, z|[x, z]\top E^R[x]|x\rangle$$

$$\Rightarrow \langle x, z|[x, z]\boxed{A, \top \vdash A}[x]|x\rangle : A$$

$$\top, A : \langle z, x|[z, x]\top E^L[x]|x\rangle$$

$$\Rightarrow \langle z, x|[z, x]\boxed{\top, A \vdash A}[x]|x\rangle : A$$

$$A : \langle x|[x]\bot I^L[z,x]|z,x\rangle$$

$$\Rightarrow \langle x|[x]\boxed{A \vdash \bot, A}[z,x]|z,x\rangle : \bot, A$$

$$A : \langle x|[x]\bot I^R[x,z]|x,z\rangle$$

$$\Rightarrow \langle x|[x]\boxed{A \vdash A, \bot}[x,z]|x,z\rangle : \bot, A$$

$$\Gamma_1, A \otimes B, \Gamma_2 : \langle V_1, z, V_2|[z] \otimes E[x,y]; [V_1, x, y, V_2]\boxed{\Gamma_1, A, B, \Gamma_2 \vdash \Delta}W|W\rangle$$

$$\Rightarrow \langle V_1, z, V_2|\boxed{\Gamma_1, A \otimes B, \Gamma_2 \vdash \Delta}W|W\rangle : \Delta$$

$$\Delta : \langle V\boxed{\Gamma \vdash \Delta_1, A, B, \Delta_2}[W_1, x, y, W_2]; [x,y] \oplus I[z]|W_1, z, W_2\rangle$$

$$\Rightarrow \langle V\boxed{\Gamma \vdash \Delta_1, A \oplus B, \Delta_2}[W_1, z, W_2]|W_1, z, W_2\rangle : \Delta_1, A \oplus B, \Delta_2$$

$$\Lambda_1, \Gamma, \Lambda_2 : \langle V_1, V, V_2|[V]\boxed{\Gamma \vdash \Delta_1, A, \Delta_2}[W_1, z, W_2];$$

$$[V_1, z, V_2]\boxed{\Lambda_1, A, \Lambda_2 \vdash \Theta}[W]|W_1, W, W_2\rangle$$

$$\Rightarrow \langle V_1, V, V_2|[V_1, V, V_2]\boxed{\Lambda_1, \Gamma, \Lambda_2 \vdash \Delta_1, \Theta, \Delta_2}[W_1, W, W_2]|W_1, W, W_2\rangle$$

$$: \Delta_1, \Theta, \Delta_2$$

where the last reduction rule, which clearly corresponds to a cut, must have at most one of each of the following pairs non-empty: Λ_1 and Δ_1, Λ_2 and Δ_2, V_1 and W_1, and V_2 and W_2. These restrictions correspond, of course, to the restrictions on the planar cut rule.

These rewrites are summarized in Fig. 1.

To extend these rules to the non-planar case we must allow for the exchange rule. This is most conveniently done by simply regarding the premise and conclusion lists of the sequent boxes as bags (multi-sets) so that their order, and the order of their corresponding ports, does not matter. This is not to suggest that the order of source or target objects is not important in the corresponding category, but merely that this approach has no effect on the property of a net being sequential. The above restriction on the cut under these assumptions becomes meaningless and so must be removed.

We should note various things about these rules: firstly the application of reduction rules must certainly terminate as every rule either "boxes" a link (or a component), has a box "eat" a link, (e.g. ($\otimes$ E) or ($\oplus$ I)), or "amalgamates" two boxes (cut) with a link. (See Fig. 1 for the graphical explanation of these terms if this is not already clear.) Thus, each reduction decreases the height in the lexicographical ordering on the number of links/components and the number of boxes.

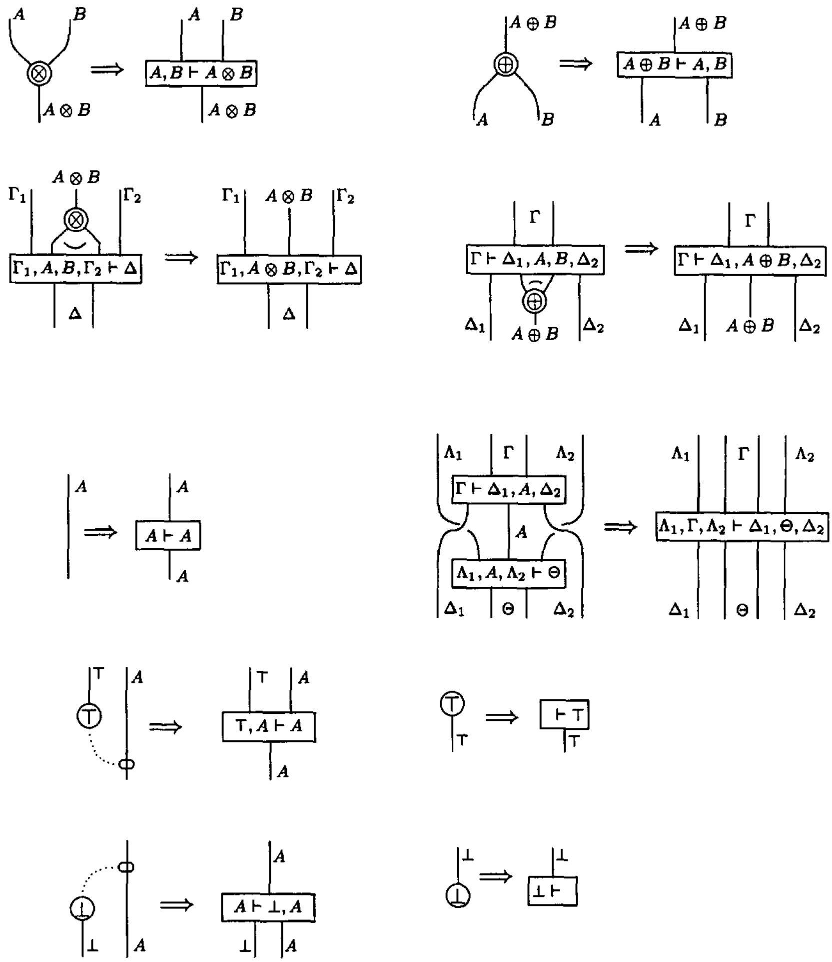


Fig. 1. Sequentialization.

Regarding the expansion rule and expansion termination: the sole expansion is easily seen to be reducing with respect to the other rules. It is also expansion promoting. This means that the system is expansion terminating provided there is a bound to the number of irreducible expansions which can be applied to a circuit. However, it is clear that one cannot expand irreducibly twice on the same wire as two such expansions introduce a trivial application of the cut rule. Thus the number of irreducible expansions is certainly bounded by the number of wires.

Finally we may ask whether the rules are confluent: as we have established that the system is reduction terminating and expansion reducing (and terminating) it suffices to show local confluence. If this is the case, the order in which this sequentialization is done will not actually matter.

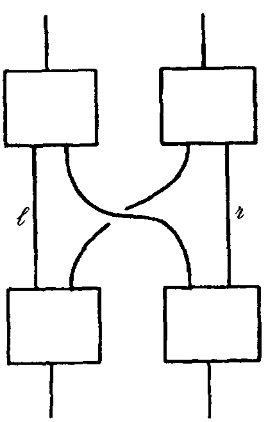
Here we must take care.

2.5. Circuit rewriting

In a rewrite system there is an obvious notion of *independent redexes* of a circuit: two redexes are independent if they can be simultaneously coalesced. Two redexes which cannot be simultaneously coalesced are called *dependent redexes*. One very obvious reason for not being able to simultaneously coalesce two redexes is that they have components in common, these are called *overlapping redexes*.

To check for confluence of the rewriting it suffices to check that all the divergences (that is pairs of overlapping rewrites) starting at dependent redexes can be resolved. This is because the independent redexes will always have an obvious resolution. In fact, one may be able to form minimal circuits exhibiting these dependencies and the divergences which result from rewriting: these are called the *critical divergences*. In standard term rewriting systems redexes are dependent if and only if they overlap; thus, the critical divergences are constructed by simply overlapping the redexes of the rules.

Unfortunately, this is not the case for circuits. The coalescing of one redex can block the formation of another redex which has no components in common. Consider the effect of applying a cut on the two vertical wires indicated:



if one applies the left-hand cut one can no longer subsequently apply the right-hand cut as the required abstraction cannot be coalesced. That is, the exchange required to bring the components together can no longer be performed. Thus, these two redexes are dependent even though they do not overlap!

The way in which a rule introduces such a non-overlapping dependence is by introducing an input/output dependence which was not originally there. A dependence between an input and output occurs when it is possible to trace a path through the circuit starting at the given input, *following the input to output direction*, and ending at the given output. In the above example, after cutting on the left, the crossing wires are forced into a new dependence as they become respectively an input and an output of the new sequent box.

A *conservative redex* is a redex which has the property that it can be simultaneously coalesced with any non-overlapping redex (in any circuit). Such a redex must have every output dependent on every input, otherwise we could introduce a cut to induce a non-overlapping dependence which would make the coalescence of that redex impossible after the cut. Conversely, if every input and output are dependent then the redex can still be coalesced after any non-overlapping redex has been coalesced. This may be seen by "growing" the coalescence inductively. The crucial step is the examination of what happens when one grows a larger conservative redex from two smaller such redexes. The only way this can happen is if the interaction completely exhausts either the input or output wires of one of the redexes. Assuming that the two smaller conservative redexes are separated by the other coalescence, the redexes whose input or output wires were exhausted by the interaction can have no interaction with an interposed redex/circuit so can be exchanged.

Lemma 2.2. *A redex is dependent on a conservative redex if and only if they overlap.*

Inspecting the rewriting rules above it is clear that all the redexes *except* the "cut" are conservative. Furthermore, it is easy to check that all the critical divergences which result from overlapping these redexes (including the cut) can be resolved. Unfortunately, of course, this is not the case for the non-overlapping dependent redexes between cuts as is illustrated above.

2.6. *Cut-cycles*

It is fortunate, therefore, that the very possibility of any such non-overlapping dependence also suffices to ensure a circuit is not sequential! The problem that remains, however, is to determine why a net cannot produce such a dependence. It is at this stage that the particular form of the sequentialization comes into play.

In the process of sequentialization, we can determine quite easily whether the only way a wire can be "removed" is as the internal wire of a cut. Inspecting the rules one concludes it must not be attached to the input port of $\oplus I$ or $\bot E$, or to an output port of $\otimes E$ or $\top I$ (as only these wires can be removed without using the cut). These wires we shall call *cut-wires*. A *cut-dependency* between two wires is present if there is a path from one wire to the other which follows the direction of all wires except the cut-wires, which are regarded as being two-way. A *cut-cycle* is a cyclic cut-dependency, that is a cycle in the circuit when the cut-wires are regarded as being two-way.

Notice that a non-overlapping dependency between cuts implies the presence of a cut-cycle. Thus, in a circuit with no cut-cycles there can be no dependencies between non-overlapping cut-redexes. In fact we may define a redex to be *cut-conservative* in case every output cut-depends upon every input.

Lemma 2.3. *Two cut-conservative redexes are dependent in a circuit with no cut-cycles if and only if they overlap.*

The proof of this is analogous to the previous proof. Cut-conservative redexes can be grown in a similar manner to conservative redexes, however, there is one additional method of growth. This is by "cutting" smaller cut-conservative redexes together: that is linking two redexes along a cut-wire such that all outputs of the upper component are cut-dependent on that wire and (dually) in the lower component the cut-wire is cut-dependent on each input.

Supposing now that the first redex is being grown by a "cutting step" and the second redex is interposed between the two parts in an effort to block the coalescence. If both parts are wired to the interposed redex then a cut-cycle is introduced. This means at most one can be so that the other can be exchanged to achieve a simultaneous coalescence.

Observe that all the redexes of sequentialization are cut-conservative. Furthermore, it is easily established, by an induction on the number of rewrites, that any circuit which can be sequentialized must certainly have no cut-cycles. This gives us:

Proposition 2.4. (i) *Any sequential circuit has no cut-cycles,*
(ii) *Sequentialization is confluent on all circuits with no cut-cycles.*

Thus, if sequentializing works one is guaranteed to end up with the same sequent. If it does not work, however, the process can become "stuck" in a number of different ways and the order of collection may, quite possibly, affect the partially collected output end result. However, whichever way it becomes stuck it will not be sequential.

Notice that to determine whether a $(\oplus, \otimes)$-circuit is sequential requires at most $n+m$ rewritings, where n is the number of components and m the number of wires of the circuit. To find an application of a rewriting, however, may also require a traversal of the structure. Thus, even from this naive view, to determine whether a circuit is sequential is an $O((n+m)^2)$ algorithm (a more detailed analysis is in [7]).

2.7. Sequentialization, the net condition, and empires

The traditional criterion for being a proof net, it is worth recalling, uses an apparently unrelated aspect of the circuits. Notice that we have drawn arcs between the $(\otimes E)$ output ports and the $(\oplus I)$ input ports: these are to indicate that these components are "switchable" in the sense introduced by Danos. This means that at most one of the ports is to be regarded as being "connected" although we do not know which way the switch is set. The classic *net criterion* on a $(\otimes, \oplus)$-circuit is satisfied if for any choice of switch settings the undirected graph determined by the wires which are judged connections is acyclic and connected. Note that the input and output wires are included in this description and are viewed as being connected to external nodes.

Following Girard [11] we have:

Proposition 2.5. *A (non-planar)* $(\oplus, \otimes)$*-circuit with components* $\mathscr{C}$ *is sequential if and only if it satisfies the net criterion.*

Proof. To see this we must first show that a sequential circuit satisfies the net criterion. This amounts to checking that each sequentialization rewriting will guarantee to box a subcircuit satisfying the net criterion. This can easily be checked by inspecting the rules.

For the converse, namely that any circuit satisfying the net criterion is sequential, we do an induction on the number of switchable components left after the sequentialization process has terminated.

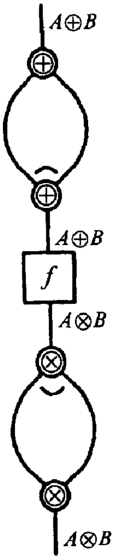
If there are none then there will be only sequent boxes left. These must however be connected in an acyclic manner by wires. Thus, if there are more than two sequent boxes left there must be a wire which is directly between sequent boxes. It is easily seen that any such wire is one along which we can perform a cut. However, by assumption we have completed sequentialization so no such wire can exist, which means there is but one sequent box and the circuit is sequential.

Assume that all circuits satisfying the net criterion, but not the proposition, sequentialize to a sequent box circuit with at least n switching components left. Considering such a sequent box circuit, select a switching component. When we disconnect its non-switching wire, the subcircuit of those components still connected to it, no matter how the switches are set, is the *(switching) empire* of that component (relative to the given disconnection).

It is easy to see that an empire satisfies the net criterion (and can be encapsulated as a subcircuit expression). Furthermore, the selected component is at the extremity of this empire and, as it is switching can be removed. This leaves a subcircuit which also satisfies the net criterion. This subcircuit has less switching components and so must be sequential. As we are assuming that sequentialization has gone as far as possible it follows that this subcircuit is just a sequent box. However, this means that the appropriate elimination rule can be applied contradicting our assumption that the net has been fully sequentialized. □

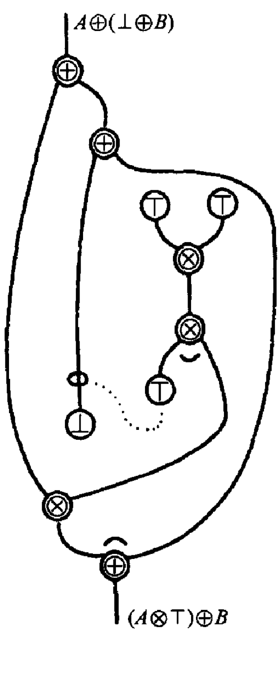
Henceforth we shall call a sequential $(\otimes, \oplus)$-circuit a *proof net* to emphasize its connection to proofs. The proposition suggests that equivalently we could have said that a circuit satisfying the net criterion is a proof net. In fact, this is the standard approach. However, there are two reasons for bucking this tradition. The first concerns the algorithmic advantage of sequentialization: naive checking of the net criterion is exponential as there are 2^n possible switch settings, where n is the number of switching components. The second, and perhaps more significant reason, is that it fails to generalize to the planar case:

Remark 2.6. (i) Troelstra [29] gives a proof of inductivity which involves searching for a "splitting" link. When non-logical axioms are present it is quite possible for there to be nets which have *no splitting links*. Consider for example the expanded normal form for an axiom such as $f : A \oplus B \longrightarrow A \otimes B$.



(ii) The proof above fails for the planar case. In particular, an empire (so defined) may not generate a subcircuit expression. In fact, it is possible to generate examples in which the switching empire "traps" some other components. This makes it impossible to apply the induction step.

(iii) Even worse, in the planar case it is quite possible for a *non-sequential circuit to satisfy the net criterion*!



(iv) The rewriting we have provided simply turns nets into logical sequents. However, it is worth mentioning that the translation should really be into the corresponding morphism of a polycategory. The critical divergences then indicate the commuting requirements of a polycategory described in [6].

The proof above introduced the concept of a switching empire. There is another way of describing this which utilizes the sequentialization process. By disconnecting a wire in a proof net we ensure that the sequentialization process cannot complete. On a sequential circuit, if we do every sequentialization step which does not "eat" the disconnected wire then to complete the sequentialization of the whole circuit the next step must be to eat that wire (unless it is at the extremity of the net). For the non-switching wire of a switching link the rule which removes it is the cut. But this means the switching link must have been absorbed into a sequent box. In fact the sequent box will have encapsulated precisely the empire of the component.

We might define a (sequential) empire of a (non-switching) component given a disconnection as the largest subnet (that is sequential subcircuit containing that component) which does not have the disconnected wire as an internal wire. Clearly this may be obtained as the "contents" of the sequent box formed by sequentializing as far as possible without eating the wire. In the commutative case this is the empire. In both the commutative and non-commutative cases this empire is important in determining to where a thinning link can be "rewired". In particular, looking ahead, the Rewiring Theorem (3.3) may be paraphrased as saying that one can rewire to anywhere on the "edge" of this empire – in the symmetric case, as every internal wire can be moved by symmetry to the edge, it follows that one can rewire to any internal wire.

3. Proof nets to weakly distributive categories

The purpose of this section is to establish the connection between proof nets, weakly distributive categories, and hence polycategories (with $\otimes$ and $\oplus$). That the last two concepts are equivalent was shown in [6]. We show that the proof nets with a single input and output form the maps of a weakly distributive category: in fact, the weakly distributive category so generated is free on the polygraph of components from which it is generated and the equations imposed. The composition of this category is given by plugging these nets together in the obvious fashion. An arbitrary proof net corresponds to a map of the two-tensor polycategory corresponding to the weakly distributive category thus generated.

In order to carry out this we must first introduce the rules of surgery for these nets. We have enumerated a more than full set of rules! [6] A minimal set is indicated and it is left to the reader to mimic the categorical techniques of Kelly [16] in diagram-

[6] We give only a generating set of the rules in the body of the text – the reader can find the other rules in Appendix A. Certainly the set of equations and rewrites may seem rather intimidating, but a glance at the graphical presentation in Fig. 2 should make things somewhat more manageable.

matic form to prove that this is indeed a minimal set. The reason for enumerating the larger set is that it is this set which forms the basis for the expansion/reduction rewriting we describe in the next section which is used to establish the coherence result.

3.1. Proof net equivalence

The equivalence of proof nets is determined by a number of surgical rules which allow one to replace one subcircuit by another. There is one assumption that we must emphasize: *a rule of surgery can only be applied to a proof net if it preserves the net criterion; that is, if after the surgical alteration, one still has a proof net.* Therefore, there is a hidden cost in applying these rules: namely one must check that the surgical alteration yields a sequential net. In fact we shall see that for most of our rules this is automatically valid, and that only those few rules that involve rewiring past a switching link do not in general preserve the net criterion and so require this extra hypothesis. More precisely, the unit rewirings given below by Eqs. (16), (17), (22), (23), (25), (26), (31), and (32) are the only surgery rules that require one to check that the right-hand side is a net if the left-hand side is.

3.1.1. Planar net equivalences

The rules of surgery for planar proof nets may be divided into the major rules of reduction and expansion followed by a myriad of rules for handling the units. While these latter are many, their pattern is rather obvious and can be explained best by diagrams which demonstrate the manipulations permitted with thinning links. These rules are not independent, however, it will be useful to have them in this form when we come to discuss, in the next section, the "rewiring" of the thinning links.

The first set of surgical rules are the *reductions*. Although we indicate a direction to these rules it is intended that they should, at the moment, be read as bidirectional. Only in the next section shall we begin to use them explicitly as rewrite rules. However, even in this section, it will be useful to the conscientious reader to know they can be used directionally, when he methodically checks the coherence diagrams for weak distributivity!

$$A,B : \langle x_1,x_2 || x_1,x_2\rangle : A,B$$

$$\Leftarrow A,B : \langle x_1,x_2 | [x_1,x_2] \otimes I[z]; [z] \otimes E[y_1,y_2] | y_1,y_2\rangle : A,B \tag{1}$$

$$A,B : \langle x_1,x_2 || x_1,x_2\rangle : A,B$$

$$\Leftarrow A,B : \langle x_1,x_2 | [x_1,x_2] \oplus I[z]; [z] \oplus E[y_1,y_2] | y_1,y_2\rangle : A,B \tag{2}$$

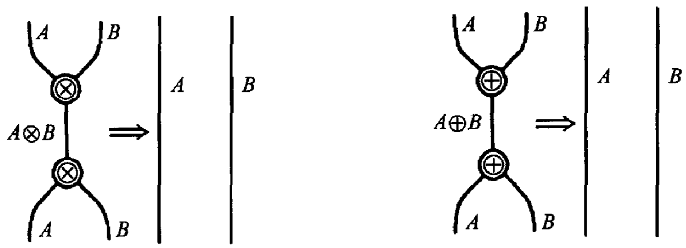
$$A : \langle x || x\rangle : A \Leftarrow A : \langle x | [\,] \top I[z]; [z,x] \top E^L[x] | x\rangle : A \tag{3}$$

$$A : \langle x||x\rangle : A \Leftarrow A : \langle x|[\,]\top I[z]; [x,z]\top E^R[x]|x\rangle : A \tag{4}$$

$$A : \langle x||x\rangle : A \Leftarrow A : \langle x|[x]\bot I^L[z,x]; [z]\bot E[\,]|x\rangle : A \tag{5}$$

$$A : \langle x||x\rangle : A \Leftarrow A : \langle x|[x]\bot I^R[x,z]; [z]\bot E[\,]|x\rangle : A \tag{6}$$

They are graphically:



(There are two dual rewrites for the units, with the unit edge and nodes on the other side of the A edge.)

The next set of surgical rules are the *expansions*. When used in the suggested rewriting direction, these should be thought of as expressing the type of the wire:

$$A\otimes B : \langle z||z\rangle : A\otimes B \Rightarrow A\otimes B : \langle z|[z]\otimes E[z_1,z_2]; [z_1,z_2]\otimes I[z]|z\rangle : A\otimes B \tag{7}$$

$$A\oplus B : \langle z||z\rangle : A\oplus B \Rightarrow A\oplus B : \langle z|[z]\oplus E[z_1,z_2]; [z_1,z_2]\oplus I[z]|z\rangle : A\oplus B \tag{8}$$

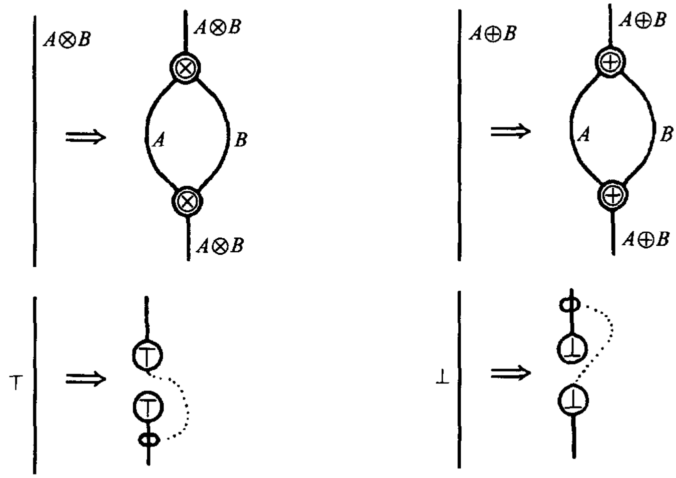
$$\top : \langle x||x\rangle : \top \Rightarrow \top : \langle x|[x]\top E^L[\,]; [\,]\top I[x]|x\rangle : \top \tag{9}$$

$$\top : \langle x||x\rangle : \top \Rightarrow \top : \langle x|[x]\top E^R[\,]; [\,]\top I[x]|x\rangle : \top \tag{10}$$

$$\bot : \langle x||x\rangle : \bot \Rightarrow \bot : \langle x|[x]\bot E[\,]; [\,]\bot I^L[x]|x\rangle : \bot \tag{11}$$

$$\bot : \langle x||x\rangle : \bot \Rightarrow \bot : \langle x|[x]\bot E[\,]; [\,]\bot I^R[x]|x\rangle : \bot \tag{12}$$

They may be displayed graphically as:



(There are two dual rewrites for the units, with the thinning edges on the other side of the unit edge and node.)

The remainder of the rules, namely those concerned with the manipulation of the thinning and cothinning links, can be found in Appendix A. Here we shall list a representative set, and shall illustrate the rules graphically in Fig. 2. (Equation numbers refer to the numbering in Appendix A.)

The rules for the unit start with the obvious ones for manipulations over the tensor, followed by less obvious rules for the interaction (as demanded by weak distributivity) of the unit with the cotensor:

$$A,\top,B : \langle x,z,y|[x,z]\top E^R[x];[x,y]\otimes I[w]|w\rangle : A\otimes B$$

$$= A,\top,B : \langle x,z,y|[z,y]\top E^L[y];[x,y]\otimes I[w]|w\rangle : A\otimes B \tag{15}$$

$$\top,A\oplus B : \langle z,x|[z,x]\top E^L[x];[x]\oplus E[x_1,x_2]|x_1,x_2\rangle : A,B$$

$$= \top,A\oplus B : \langle z,x|[x]\oplus E[x_1,x_2];[z,x_1]\top E^L[x_1]|x_1,x_2\rangle : A,B \tag{20}$$

$$A\oplus B,\top : \langle x,z|[x,z]\top E^R[x];[x]\oplus E[x_1,x_2]|x_1,x_2\rangle : A,B$$

$$= A\oplus B,\top : \langle x,z|[x]\oplus E[x_1,x_2];[x_2,z]\top E^R[x_2]|x_1,x_2\rangle : A,B \tag{21}$$

$$A\oplus B : \langle w|[w]\oplus E[x,y];[x]\perp E^R[x,z];|x,z,y\rangle : A,\perp,B$$

$$= A\oplus B : \langle w|[w]\oplus E[x,y];[y]\perp E^L[z,y];|x,z,y\rangle : A,\perp,B \tag{24}$$

$$A,B : \langle x_1,x_2|[x_1,x_2]\otimes I[x];[x]\perp I^R[x,z]|x,z\rangle : A\otimes B,\perp$$

$$= A,B : \langle x_1,x_2|[x_2]\perp I^R[x_2,z];[x_1,x_2]\otimes I[x]|x,z\rangle : A\otimes B,\perp \tag{29}$$

$$A,B : \langle x_1,x_2|[x_1,x_2] \otimes I[x]; [x]\bot I^L[z,x]|x,z\rangle : \bot, A \otimes B$$

$$= A,B : \langle x_1,x_2|[x_1]\bot I^L[z,x_1]; [x_1,x_2] \otimes I[x]|x,z\rangle : \bot, A \otimes B \tag{30}$$

$$\Gamma_1,A,\top,B,\Gamma_2 : \langle ..,x_1,z,x_2,..|[x_1,z]\top E^R[x_1]; [..,x_1,x_2,..]f[..]|..\rangle : \Delta$$

$$= \Gamma_1,A,\top,B,\Gamma_2 : \langle ..,x_1,z,x_2,..|[z,x_2]\top E^L[x_2]; [..,x_1,x_2,..]f[..]|..\rangle : \Delta \tag{49}$$

$$\Gamma,A,\top : \langle ..,x_1,z|[x_1,z]\top E^R[x_1]; [..,x_1]f[..,x_2]|..,x_2\rangle : \Delta,B$$

$$= \Gamma,A,\top : \langle ..,x_1,z|[..,x_1]f[..,x_2]; [x_2,z]\top E^R[x_2]|..,x_2\rangle : \Delta,B \tag{51}$$

$$\top,A : \langle z,x|[z,x]\top E^L[x]|x\rangle : A$$

$$= \top,A : \langle z,x|[x,z]\top E^R[x]|x\rangle : A \tag{57}$$

$$\Gamma,\top : \langle ..,z|[..]f[..,x_1,x_2,..]; [x_1,z]\top E^R[x_1]|..,x_1,x_2,..\rangle : \Delta_1,A,B,\Delta_2$$

$$= \Gamma,\top : \langle ..|[..]f[..,x_1,x_2,..]; [x_2,z]\top E^L[x_2]|..,x_1,x_2,..\rangle : \Delta_1,A,B,\Delta_2 \tag{59}$$

These are displayed graphically in Fig. 2. Note that Eq. (23) (marked with a $*$) and the other rewirings past switching links, (16), (17), (22), (25), (26), (31), (32), are the only rewirings that require one to check that the right-hand side is a net if the left-hand side is a net. In the other equations one can easily check that the net condition is preserved.

3.1.2. Surgery, sequentiality, and dependence

It is worth recalling at this stage that many of the rules, if driven in the correct direction, will preserve the property of being a proof net whenever they are applied. Notice, in particular, that the reduction rules, when driven in the indicated direction, certainly will preserve sequentiality. However, if these rules are driven in the wrong direction, there is no guarantee that sequentiality will be preserved. The expansion rules preserve sequentiality in both directions.

All the thinning link rules involving other components which are non-switching (in particular other thinning links) are of this form. However, the rules involving switching components can, if misapplied, result in a circuit which is not sequential. There are exceptions to this: for example the rules involving the unit thinning and the tensor elimination (and dually the counit thinning and the cotensor introduction) will preserve sequentiality when used in either direction.

As mentioned earlier this is a highly redundant set of rules. In particular, all the unit and counit manipulations rules (15)–(48) can be reduced to just six rules: (15), (20), (21), (24), (29), and (30). This is a challenging exercise for the reader! It is useful to realize that many of these follow quite easily from the coherence of the tensors and the tricks in [16].

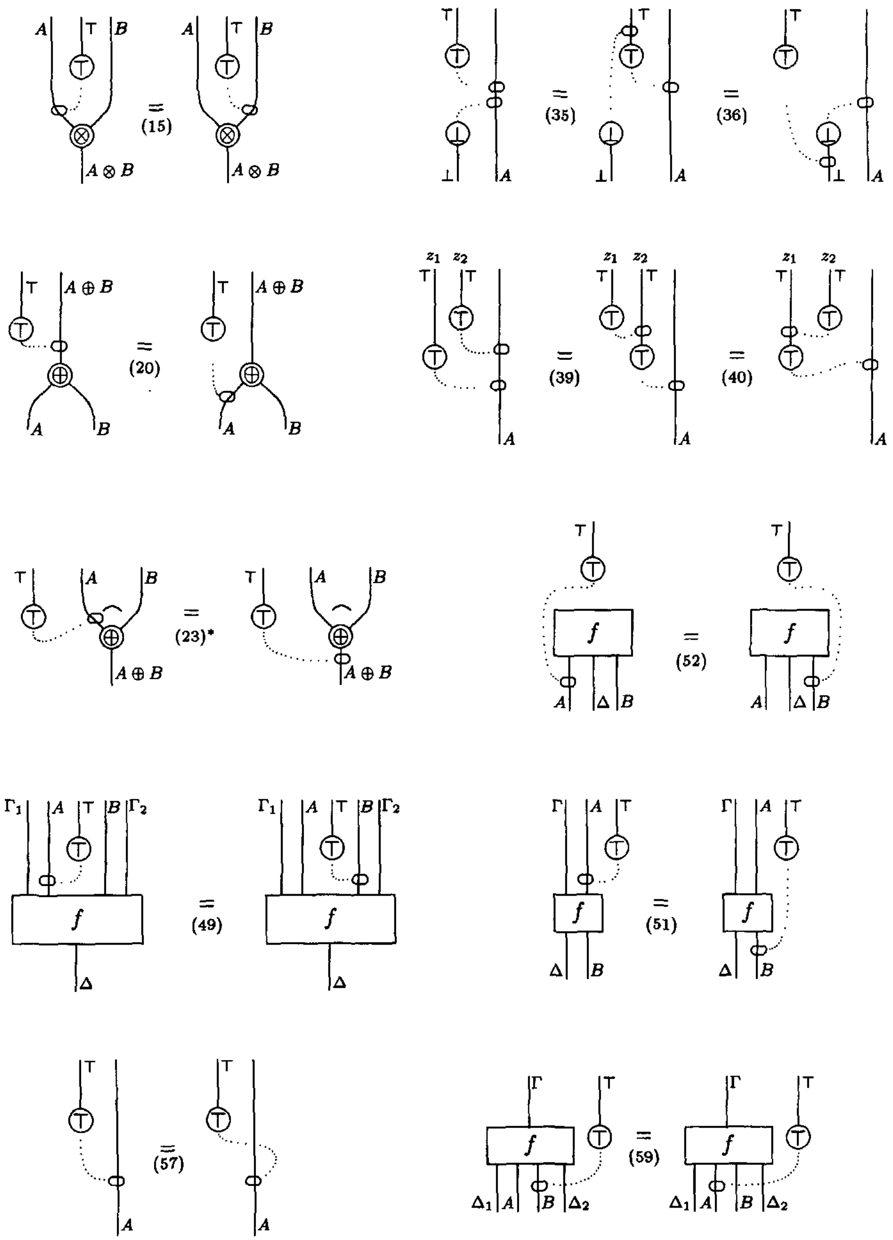
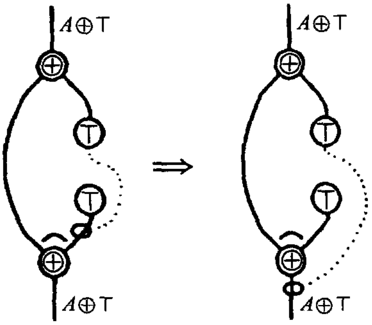
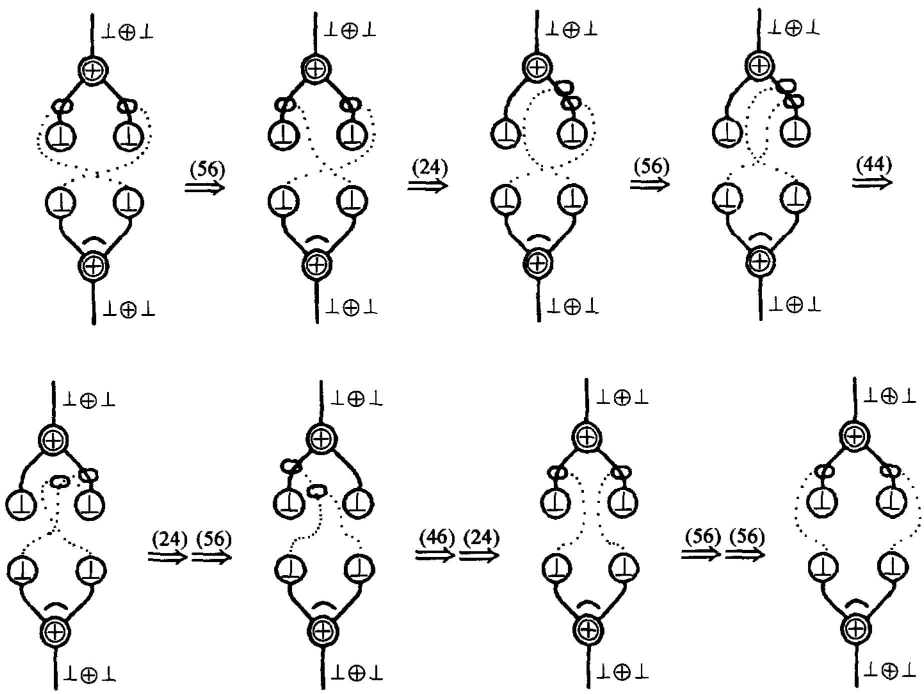


Fig. 2. Unit rewirings.

We end with some examples of the unit rewiring steps. First an illustration of an illegal application of Eq. (22); it is easy to check that the net criterion is not preserved in the following rewrite.



To illustrate a valid use of the rewirings, note how the following uses of Eqs. (56), (46), (44), and (24) transform the "twist" map $\perp \oplus \perp \longrightarrow \perp \oplus \perp$ into the identity. We leave it to the reader to verify that a similar set of steps does not exist to transform the twist map $\top \oplus \top \longrightarrow \top \oplus \top$ into the identity; this is as expected, since these maps are not equal, whereas the other two are, in the free weakly distributive category. Once we have the Rewiring Theorem, it will be easy to show that in fact no rewiring can transform the twist on $\top \oplus \top$ into the identity because the unit empires are all trivial.

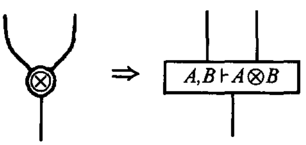


3.2. The Rewiring Theorems

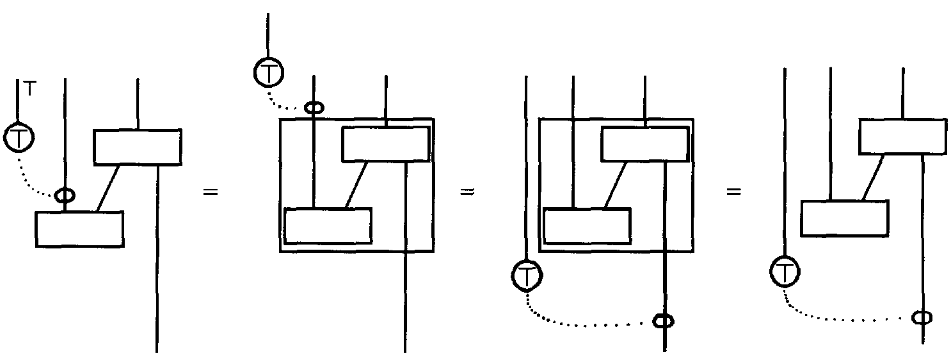
In this section we show that the "box rewiring" rules (Eqs. (49)–(56)) can be extended so that the boxes (which originally represented components) may be interpreted as arbitrary subnets: any subnet which can be collected using the sequentialization process can be treated by the units as if it were a primitive component for the purposes of these rules. This allows us to move thinning links in large steps which we will call "empire moves" as in the symmetric case these are movements to wires in the same empire. This derived rule is quite powerful, and lies at the heart of our coherence results. Rewiring works in both the non-symmetric and symmetric case. However, there is a significant difference between the two cases. In the non-symmetric case we have to use the rules for eliminating units, Eqs. (3)–(6), as two-way rules in order to achieve an arbitrary rewiring. In the symmetric case we can avoid using these rules by utilizing (57), (58). This results in the following two versions of the rewiring theorem:

Proposition 3.1 (Non-commutative Rewiring Theorem). *The box rewiring rules* (*Eqs.* (49)–(56)) *apply to any subnet of a planar net. Furthermore, in order to achieve these rewirings only Eqs.* (3)–(6) *and* (15)–(48) *need be used.*

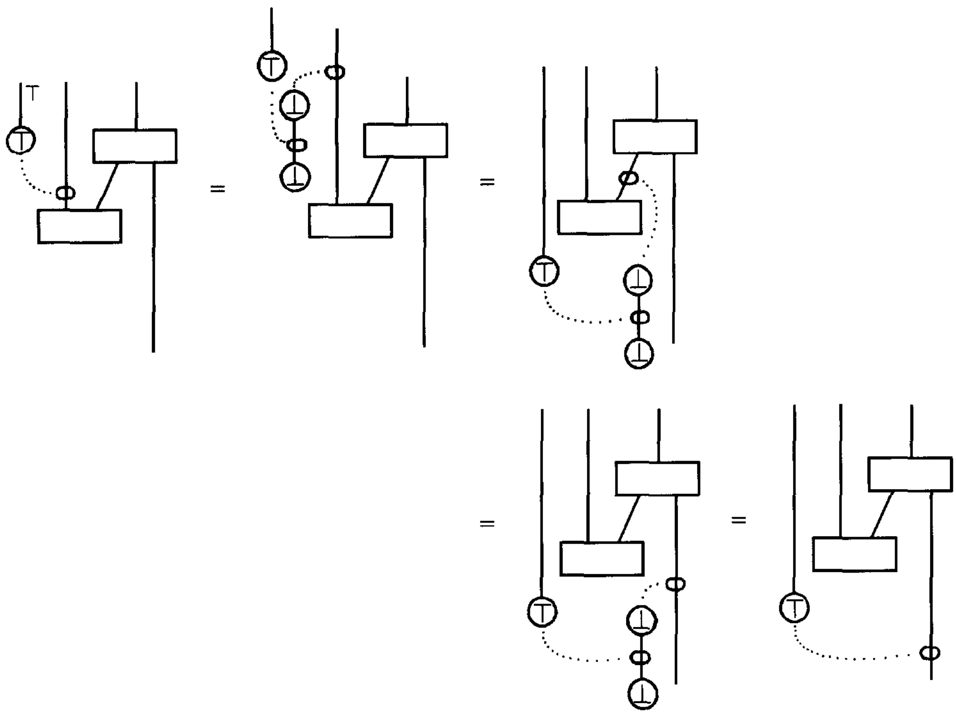
Proof. We need to check that each sequentialization rule creates box rewiring moves. For example, consider the sequentialization



We must check that the implied box rewirings for this box (that is, the rewirings that would be implied if the box were to be considered as a component) are already possible for the tensor introduction link. But this is trivially the case. Similarly it is necessary to check each sequentialization rule (in the symmetric and non-symmetric case). Most of this is straightforward; the family of cases which require some special argument arise from special cases of the cut rule where a subnet has either no inputs or no outputs. For example:



Our problem is to accomplish this rewiring by an induction. In the commutative case this is easy – we could just move the thinning link from one input/output wire to another using Eqs. (49)–(56), (59), (60). In the non-commutative case this is a very different matter. The problem is that without Eqs. (59), (60), our rules do not allow moving a $\top$ thinning link around the bottom of the box without outputs, but rather do allow one to so move a $\perp$ thinning link. So the trick is to use a "floating bar-bell" which can eventually be eliminated to carry the $\top$ thinning link across. A dual arrangement also applies for moving a $\perp$ thinning link around the top of a box without inputs.

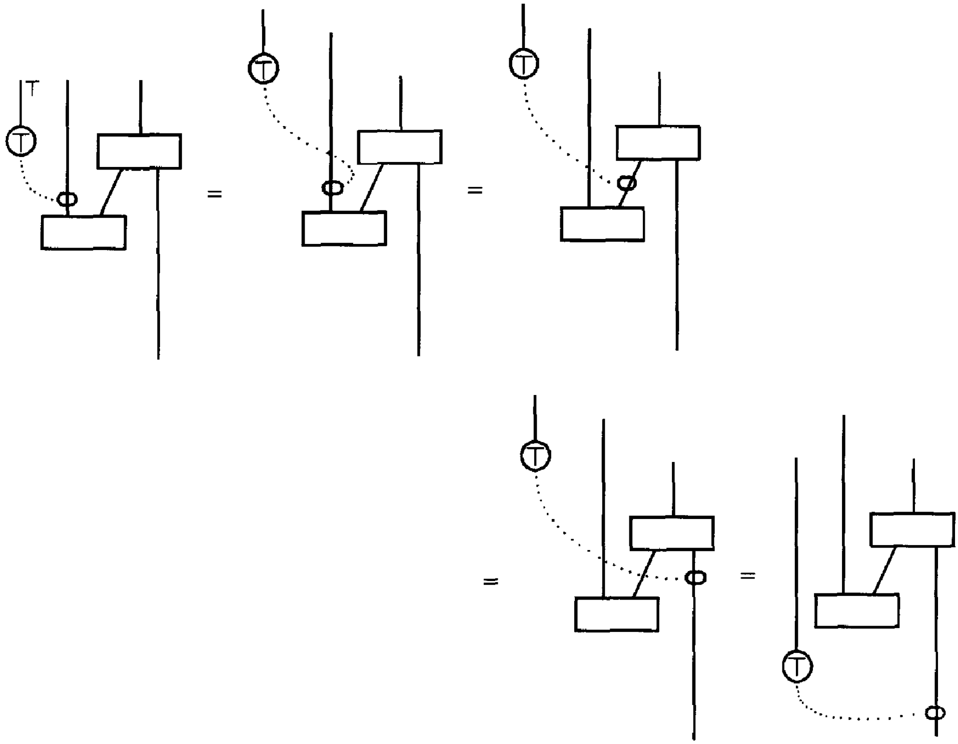


The use of a "reverse reduction" here is valid – the net criterion is preserved, and it is clear that the net in the second figure above is indeed equal to the net in the first. But this is why we need to add the unit reductions Eqs. (3)–(6) to the set of equations.

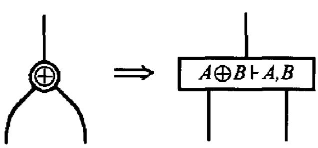
□

Proposition 3.2 (Commutative Rewiring Theorem). *The box rewiring rules (Eqs.* (49)–(60)) *apply to any subnet of a non-planar net. Furthermore, in order to achieve these rewirings only Eqs.* (15)–(48) *and* (57)–(58) *need be used.*

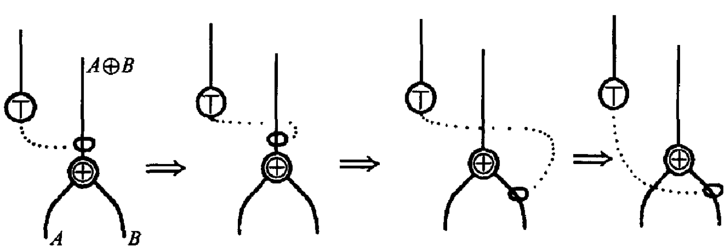
Proof. The proof above goes through in the symmetric case, with one simplification: the need for the "bar-bell" move is gone (and hence we no longer need to add the reverse unit reductions to the equations). Compare the following moves with the "bar-bell" moves above:



In the symmetric case we also must check Eq. (59) (Eq. (60) is dual); this amounts to checking the following box:



But this is exemplified by the following:



□

In [28] a proof of what is essentially Proposition 3.2 above was given. The proof was quite different and applied only to the symmetric case. It used the notion of empire to determine the wires to which a thinning link could be moved. As these are fairly standard notions it provides a useful alternative view of the rewiring theorem in the commutative case.

In the commutative case the order and manner in which thinning links are attached to a wire is not important, as Eqs. (57), (58) allow one to flip and hence interchange the order. Furthermore, what local movements of a thinning link are possible is no longer affected by what sort of link it is. In fact, the only restriction on a local move is that the net criterion (given by switchings) is preserved by the move. This allows a thinning link to be moved from its initial "home" wire to any wire which is still connected to that home wire, whatever the switch settings, when the thinning link itself is not used for that connection. This is precisely what is meant by the empire of the link (see [2] for more details on empires). Thus, we have:

Proposition 3.3 (Empire rewiring). *In a non-planar net a thinning link can be moved to any wire in its empire.*

Proof. "Only if" is a consequence of the fact that the rewiring steps defined in Section 3 are all within the appropriate unit's empire. "If" is essentially the content of Proposition 3.2, since the empire of a formula is a subnet. □

Note that if a wire is in the empire of the thinning link then there must be a subnet having these wires as inputs or outputs, comprising those components connected to both while not using those wires. Thus, Propositions 3.2 and 3.3 are equivalent. One further observation: in the free symmetric case the empire of the thinning link is also the largest set of wires to which a thinning link can be moved while preserving the Lambek equivalence of proofs.

3.3. Proof nets as weakly distributive categories

A proof net with one input and one output can be regarded as a morphism between the type of its input wire and the type of its output wire. Clearly such proof nets can be plugged together: sequentiality is preserved as each may be sequentialized and then the two can be cut together. The fact that composition is associative follows from the associativity of the "plugging together" operation (juxtaposition of circuits). The identities are the single wires. Notice even before we consider the equivalences we now have a category. Finally, the equivalences define a congruence on this category: if $n_1 \equiv n_1'$ and $n_2 \equiv n_2'$ then, as the surgeries for obtaining the two equivalences are disjoint, $n_1; n_2 \equiv n_1'; n_2'$. Furthermore we can add arbitrary equivalences between proof nets having components and this result will still be true. The fact that both sides of such an equivalence must be nets based on the same sequence of inputs and outputs implies that no new switching links can be added and that the equivalence will preserve sequentiality when applied in either direction.

Thus, proof nets based on some set of components $\mathscr{C}$ quotiented by the equivalences above and any set E of equivalences form a category $\mathrm{Net}_E(\mathscr{C})$ (and $\mathrm{PNet}_E(\mathscr{C})$) for the planar nets). The point, of course, is that:

Proposition 3.4. $\mathrm{Net}_E(\mathscr{C})$ *is a symmetric weakly distributive category and* $\mathrm{PNet}_E(\mathscr{C})$ *is a (non-symmetric) weakly distributive category.*

The translation of the components of a weakly distributive category into proof nets is as follows:

$$f \otimes g = [x] \otimes E[x_1, x_2]; [x_1] f [y_1]; [x_2] g [y_2]; [y_1, y_2] \otimes I[y]$$

$$f \oplus g = [x] \oplus E[x_1, x_2]; [x_1] f [y_1]; [x_2] g [y_2]; [y_1, y_2] \oplus I[y]$$

$$a_\otimes = [x] \otimes E[x_1, x_2]; [x_1] \otimes E[y_1, z]; [z, x_2] \otimes I[y_2]; [y_1, y_2] \otimes I[y]$$

$$(a_\otimes)^{-1} = [x] \otimes E[x_1, x_2]; [x_2] \otimes E[z, y_2]; [x_1, z] \otimes I[y_1]; [y_1, y_2] \otimes I[y]$$

$$a_\oplus = [x] \oplus E[x_1, x_2]; [x_1] \oplus E[y_1, z]; [z, x_2] \oplus I[y_2]; [y_1, y_2] \oplus I[y]$$

$$(a_\oplus)^{-1} = [x] \otimes E[x_1, x_2]; [x_2] \oplus E[z, y_2]; [x_1, z] \otimes I[y_1]; [y_1, y_2] \oplus I[y]$$

$$\delta_L^L = [x] \otimes E[x_1, x_2]; [x_2] \oplus E[z, y_2]; [x_2, z] \otimes I[y_1]; [y_1, y_2] \oplus I[y]$$

$$\delta_R^R = [x] \otimes E[x_1, x_2]; [x_1] \oplus E[y_1, z]; [z, x_1] \otimes I[y_2]; [y_1, y_2] \oplus I[y]$$

$$u_\otimes^L = [x] \otimes E[z, x]; [z, x] \top E^L[y]$$

$$(u_\otimes^L)^{-1} = [\,] \top I[z]; [z, x] \otimes I[y]$$

$$u_\otimes^R = [x] \otimes E[x, z]; [x, z] \top E^R[y]$$

$$(u_\otimes^R)^{-1} = [\,] \top I[z]; [x, z] \otimes I[y]$$

$$u_\oplus^L = [x] \otimes E[z, y]; [z] \bot E[\,]$$

$$(u_\oplus^L)^{-1} = [x] \bot E^L[z, x]; [z, x] \oplus E[y]$$

$$u_\oplus^R = [x] \otimes E[y, z]; [z] \bot E[\,]$$

$$(u_\oplus^R)^{-1} = [x] \bot E^R[x, z]; [x, z] \oplus E[y]$$

$$s_\otimes = [x] \otimes E[z_1, z_2]; [z_2, z_1] \otimes I[y]$$

$$s_\oplus = [x] \oplus E[z_1, z_2]; [z_2, z_1] \oplus I[y]$$

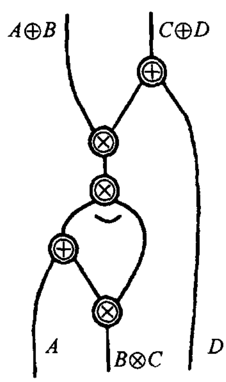
It is necessary to show that $\_\otimes\_$ and $\_\oplus\_$ are functors, and that the weak distributions, and associativities (and symmetries when present), are natural isomorphisms. The proof of all these is straightforward and is left to the reader. We also have to show that the unit and counit introduction and elimination maps are natural. That the unit introduction map $(u_\otimes^R)^{-1}$ is natural is easily established as is the fact that unit introduction and elimination (thinning) are inverse. Thus, unit elimination (thinning) is natural as it is the inverse of a natural transformation. A similar argument works for the counit introduction (cothinning).

The proof of the theorem then involves checking that the diagrams described in [6] all commute. Checking the diagrams is, in fact, absolutely straightforward: we demonstrate the idea with some examples.

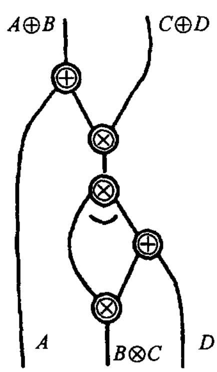
Consider the proof nets that arise from the two composites in the following diagram (one of the "coherence conditions" for weakly distributive categories [6]):

$$\begin{tikzcd}
 & (A\oplus B)\otimes(C\oplus D) \arrow[dl, "\delta^L_L"'] \arrow[dr, "\delta^R_R"] & \\
((A\oplus B)\otimes C)\oplus D \arrow[d, "\delta^R_R\oplus i_D"'] & = & A\oplus(B\otimes(C\oplus D)) \arrow[d, "i_A\oplus\delta^L_L"] \\
(A\oplus(B\otimes C))\oplus D \arrow[rr, "a_\oplus"] & & A\oplus((B\otimes C)\oplus D)
\end{tikzcd}$$

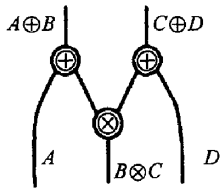
Going around the left-hand side gives a net that partially reduces to



And going around the right-hand side gives a net that partially reduces to



It is easy to see that each of these reduces to

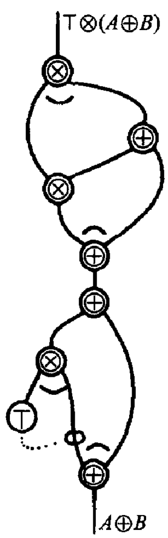


So the diagram commutes, as the two composites can be surgically altered to be the same net. Notice that we have suppressed the tensors in the source and the cotensors in the target to simplify these nets; the appropriate links at the top and bottom can be added to obtain the morphisms.

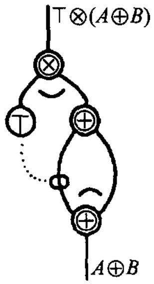
Similarly, consider the diagram

$$\begin{tikzcd}
\top \otimes (A \oplus B) \arrow[d, "\delta_L"'] \arrow[dr, "u^L_\oplus"] & \\
(\top \otimes A) \oplus B \arrow[r, "u^L_\otimes \oplus i_B"'] & A \oplus B
\end{tikzcd}$$

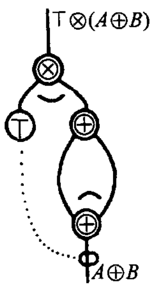
The composite map is, when expressed as a net,



which easily reduces to



The rewiring move given by Eq. (23) shows this equal to the following, which is the expanded normal form of $u^L_\otimes$. So the diagram commutes.



We claim in fact not only that these nets give a weakly distributive category but also that they provide the free such categories generated by the polygraph $\mathscr{C}$ of components and the equivalences E.

Theorem 3.5. $\mathrm{Net}_E(\mathscr{C})$ *is the free symmetric weakly distributive category generated by the polygraph* $\mathscr{C}$ *and the equations* E*. Similarly,* $\mathrm{PNet}_E(\mathscr{C})$ *is the free (non-symmetric) weakly distributive category generated by this data.*

Proof. It suffices to show that all the rules of surgery are consequences of the axioms of weak distributivity. This can be done best by translating the rules back into morphisms and checking that indeed the diagrams in question must commute. To do this one must complete the surgical rules to morphisms and make explicit the tensor/cotensor arrangement of the free wires.

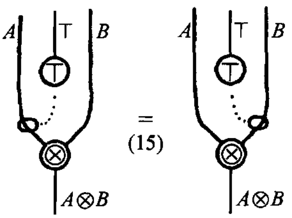
The reductions and expansions for the tensor and cotensor are fairly obviously valid – indeed they are essentially the polycategorical bijections

$$\frac{\Gamma \longrightarrow \Delta}{\bigotimes \Gamma \longrightarrow \Delta} \qquad\qquad \frac{\Gamma \longrightarrow \Delta}{\Gamma \longrightarrow \bigoplus \Delta}$$

And similarly the reductions and expansions for the units are essentially the polycategorical bijections

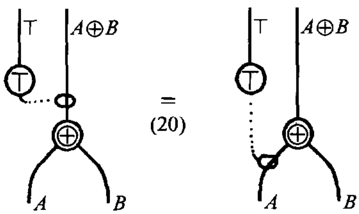
$$\frac{\Gamma,\Gamma' \longrightarrow \Delta}{\Gamma,\top,\Gamma' \longrightarrow \Delta} \qquad\qquad \frac{\Gamma \longrightarrow \Delta,\Delta'}{\Gamma \longrightarrow \Delta,\bot,\Delta'}$$

So we really only need to illustrate the rewirings of thinning links. Here are a few; we leave the rest to the reader.



amounts to

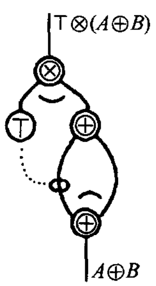
$$\begin{tikzcd}
(A\otimes\top)\otimes B \arrow[rr, "a_\otimes"] \arrow[dr, "u^R_\otimes \oplus i_B"'] & & A\otimes(\top\otimes B) \arrow[dl, "i_A \oplus u^L_\otimes"] \\
& A\otimes B &
\end{tikzcd}$$



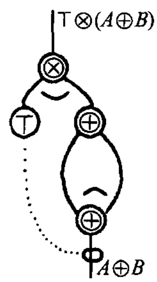
amounts to

$$\begin{tikzcd}
\top\otimes(A\oplus B) \arrow[d, "\delta^L_L"'] \arrow[dr, "u^L_\otimes"] & \\
(\top\otimes A)\oplus B \arrow[r, "u^L_\otimes \oplus i_B"'] & A\oplus B
\end{tikzcd}$$

These are both among the basic commuting diagrams imposed on weakly distributive categories by [6]. The first is one of the basic coherence conditions for monoidal categories, and the second is one of the diagrams imposed connecting the units with the weak distributivities. This is sufficient to illustrate the generating equations in fact. As an illustration of Eq. (23) we note that the condition that the rewrite preserves the net criterion implies that (possibly after some expansion) the part of the net containing the figure concerned must look something like the following:



which by Eq. (23)* should =

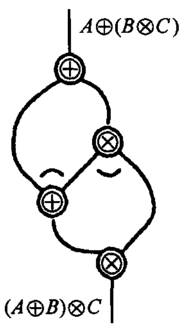


and this is essentially the same as

$$\begin{tikzcd}
\top \otimes (A \oplus B) \arrow[d, "\delta_L^L"'] \arrow[dr, "u_\otimes^L"] & \\
(\top \otimes A) \oplus B \arrow[r, "u_\otimes^L \oplus i_B"'] & A \oplus B
\end{tikzcd}$$

as we have already seen. □

Using the net criterion for proof nets, it is easy to tell if a given hom-set in the free category is inhabited. Given any two objects A, B, there are finitely many proof structures that could be associated with the sequent $A \longrightarrow B$; it is straightforward to check if any satisfy the net criterion. For example, there is in general no "dual distributivity" map $A \oplus (B \otimes C) \longrightarrow (A \oplus B) \otimes C$; the associated proof structure in "expanded normal form" would be



which clearly does not satisfy the net criterion.

4. Coherence

In this section we address the question of when two morphisms in a free weakly distributive category are equal. The crucial step in this has now been accomplished: namely the presentation, given by Theorem 3.5 above, of weakly distributive categories as nets. In the symmetric case we shall completely settle this question. In the non-symmetric case, while we considerably narrow the problem, we shall leave it open: a further analysis of the residual non-symmetric rewirings is required which is beyond the scope of this paper.

A "free" weakly distributive category is one generated from a polygraph $\mathscr{C}$ with no additional net equivalences: that is $\mathrm{Net}_{\emptyset}(\mathscr{C})$. To provide a decision procedure for these nets we regard the basic net equivalences as an expansion/reduction system modulo equations: this organization was introduced in Section 3.1. However, it remains to establish that the requirements, as laid out in Appendix B, for a strongly normalizing expansion/reduction system modulo equations are satisfied. The confirmation of this constitutes the main technical result of this section.

In the symmetric case the residual equations are the rewirings, and since a net has only a finite number of possible rewirings, the expansion/reduction rewriting yields a decision procedure. In the non-symmetric case it is necessary to include the unit reductions as two-way rules and this makes a decision procedure for the residual equations less easy to obtain; indeed, as mentioned above, we have left this open.

We end the section by examining a famous example in the coherence theory of monoidal categories: the "triple-dual" problem. It illustrates how the rewirings allow one to characterize equality of morphisms in that case. It also is a reminder of the non-trivial role of the rewiring equations.

4.1. The expansion/reduction system for nets

Using the organization of the equivalences suggested in Section 3.1 we can state the main technical theorem of this section:

Theorem 4.1 (Coherence: commutative case). *Let $\mathscr{C}$ be any polygraph; then for $\mathrm{Net}_{\emptyset}(\mathscr{C})$ the system of tensor/cotensor reductions and expansions modulo the rewirings of thinning links is an expansion-terminating expansion/reduction system modulo equations (in the sense of Appendix B).*

Note this implies uniqueness of expanded normal forms modulo the equivalences given by the rewirings.

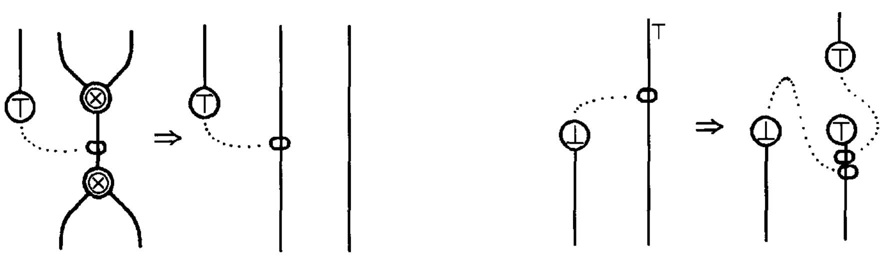
Proof. See Appendix B for definitions and terminology. We use Theorem B.6. We denote by $\mathscr{R}$, $\mathscr{X}$, and $\mathscr{E}$ the reductions, expansions, and equivalences of $\mathrm{Net}_{\emptyset}(\mathscr{C})$. Notice first that:

- The reduction rules are $\mathscr{E}$-terminating, since they all eliminate links, whereas no $\mathscr{E}$-rule does this.
- The expansion rules are expansion terminating, because the system is expansion promoting (so that rewrites may be rearranged so as to have expansions precede reductions) and hence the number of irreducible expansions on a net is bounded by the number of wires.
- So what must be proved is that $\mathscr{X} \cup \mathscr{R}$ is $\mathscr{X}$-reducing and locally $\mathscr{E}$-confluent: given a $\mathscr{X} \cup \mathscr{R}$ divergence $n_1' \xrightarrow{v_1} n_1 \underset{\sim}{\xrightarrow{e^*}} n_2 \longrightarrow^{v_2} n_2'$ there is an $\mathscr{X}$-reducing convergence modulo $\mathscr{E}$.

In the symmetric case there is a simplification in the handling of the unit rewirings that we have already seen in the proof of Proposition 3.2, namely that we only need pay attention to what wires the thinning links are attached to, and not to the manner or order in which they are attached, since they can be interchanged. Indeed, in determining whether a rewiring is possible, we do not have to worry about which type of unit is involved.

Definition 4.2 (*skeleton*). The skeleton of a net is the graph obtained from the net by removing all thinning links.

Reductions and expansions of skeletons are clearly confluent. Given a reduction (or expansion) v we define a "pseudo-reduction (or expansion)", that is a mapping $sk[v]$ of the thinning link structure, which mimics $\mathscr{E}$-moves on the original net with corresponding moves of the mapped thinning link structure on the reduced (or expanded) skeleton. (We are not claiming – at this stage – the equivalence of these nets.) We define $sk[v]$ for basic skeletal reductions and expansions by considering where thinning links might be attached – note that if a thinning link is not attached to any wire involved in the reduction or expansion, there is nothing to do. For example, in the cases $v = \otimes$-reduction, $v = \top$-expansion, if a thinning link were to be attached to the primary wire (that is, the "middle" or $A \otimes B$ wire, and the $\top$ wire respectively), then $sk[v]$ would be



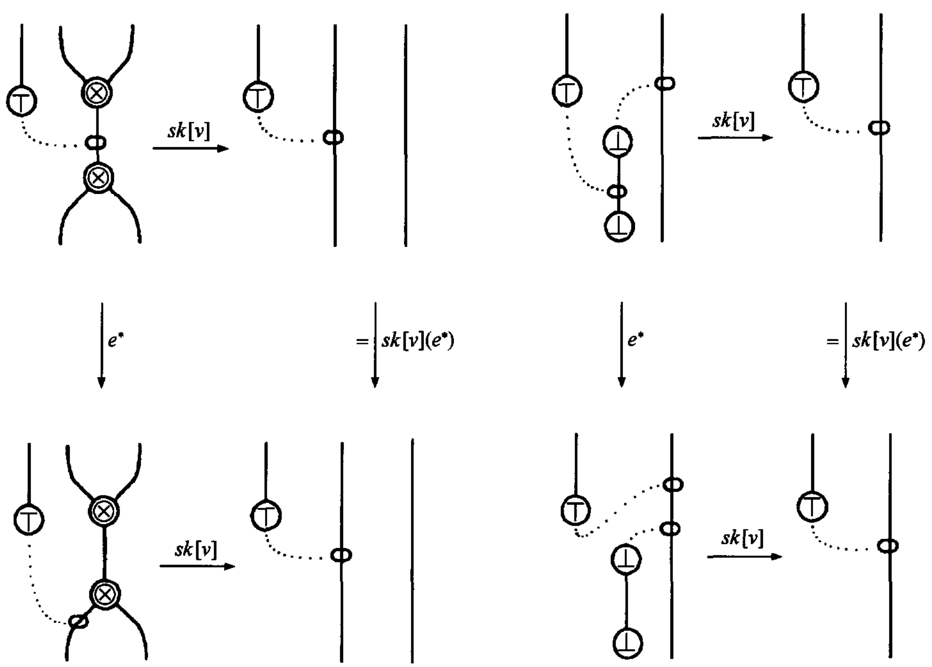
We leave the other cases to the reader. Some remarks: First note that a reduction or expansion such as those above is strictly speaking not allowed by our presentation, as our reduction rules did not account for the presence of a thinning link on the primary wire (a thinning link elsewhere could be ignored since the thinning link in such a case does not overlap the reduction or expansion). The point of the skeleton map is that it allows us to do such reductions and expansions and carry the thinning link along as well. This type of move is of course equivalent to first moving the thinning link out of the way via the evident rewiring and then performing the reduction, which will bring us to our next definition. Next, it is clear that there is a degree of arbitrariness in the choice we have made by moving the thinning link – in the symmetric case we could have moved it to the right wire instead of to the left wire, or in the case of the unit expansion, to above instead of below the expanded unit wire – but we have tried to always make the choice that is compatible with the non-symmetric case. For our present purposes, however, it really does not matter what recipe we give as long as it is definite.

So for arbitrary $n \xrightarrow{v} n' \in (\mathscr{R} \cup \mathscr{X})^*$ we have defined a map $n \xrightarrow{sk[v]} n'$. Now we extend this to the $\mathscr{E}$-moves: given $n_1 \xrightarrow{v} n_1' \in (\mathscr{R} \cup \mathscr{X})^*$ and $n_1 \xrightarrow{e^*} n_2 \in \mathscr{E}$ we define an $\mathscr{E}$-move $sk[v](e^*)$.

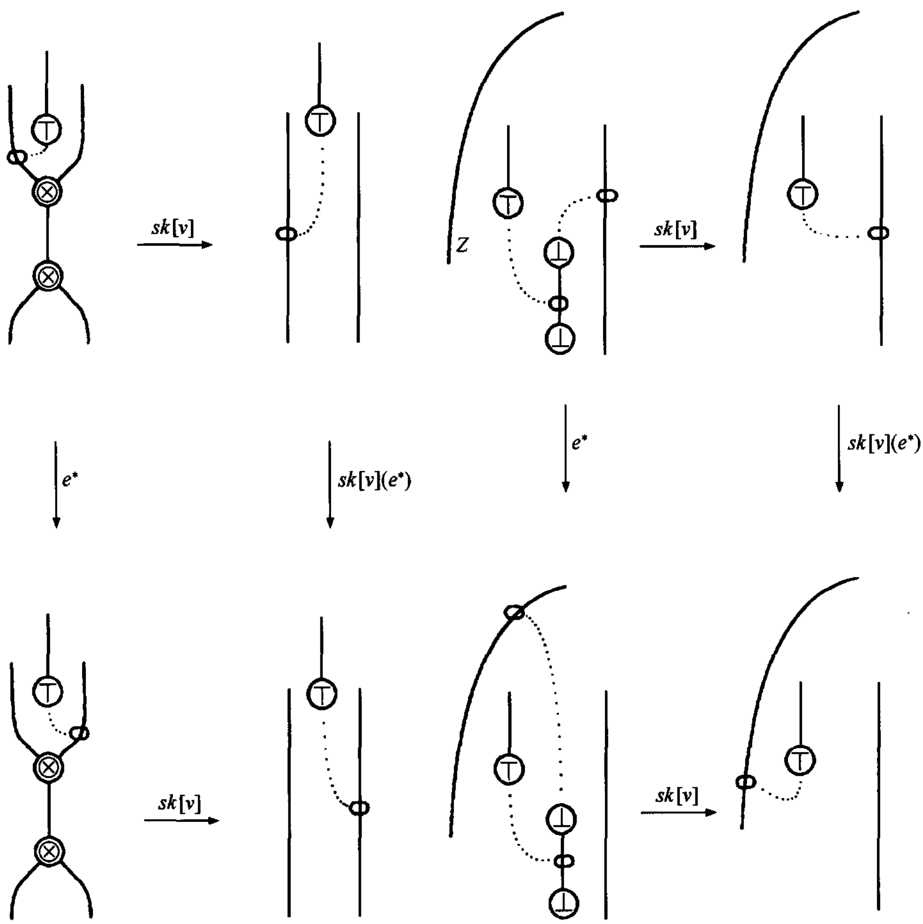
$$
\begin{tikzcd}
n_1 \arrow[r, "sk[v]"] \arrow[d, "e^*"'] & n'_1 \arrow[d, "sk[v](e^*)"] \\
n_2 \arrow[r, "sk[v]"] & n'_2
\end{tikzcd}
$$

Note that since $sk[v]$ is defined on the skeleton, which is not affected by e^*, it is indeed the same $sk[v]$ that appears as $n_1 \longrightarrow n_1'$ and $n_2 \longrightarrow n_2'$.

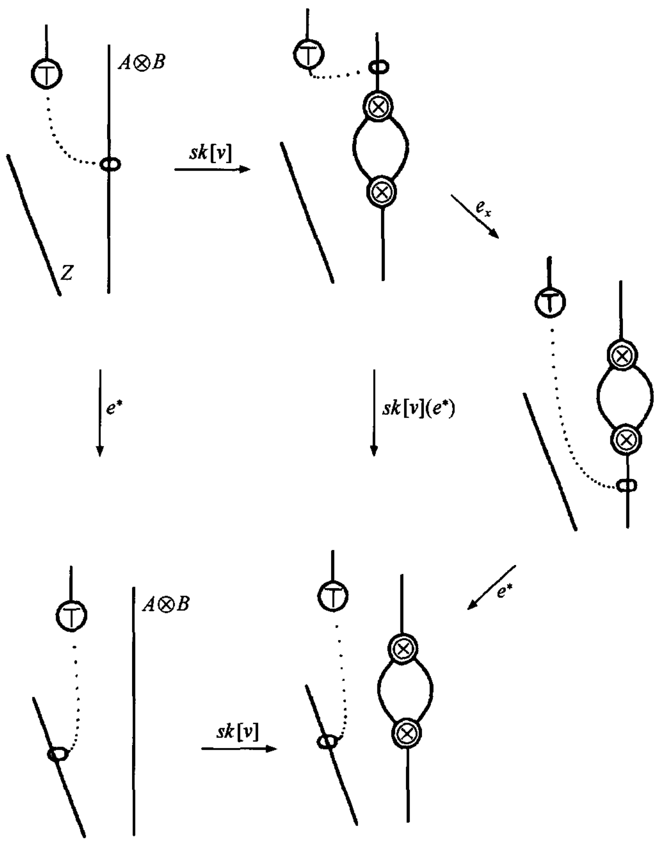
This map is given by describing how to handle the $\mathscr{E}$-moves which overlap the reductions/expansions (others are left unaltered – i.e. the mapping is the identity on such). For example, here are two reduction cases where $sk[v](e)$ is the identity.



The following are two examples where $sk[v](e^*)$ is more complicated, being given by a "box-rewiring" or empire move corresponding to moving between the input wires of a subnet which must exist since the configuration shown is part of a net and so satisfies the net criterion. In particular, on the left-hand side the "parallel" wires shown must in fact be connected by some such subnet. And on the right-hand side, the $\top$ rewiring given by $sk[v](e^*)$ just mimics the $\perp$ rewiring given by e^*. In the commutative case this is possible, since we do not have to worry about wire-crossings and since (as mentioned before) we do not have to worry about which unit is involved to decide if a rewiring is possible. In the non-commutative case this is not possible, and so this stage of our proof cannot proceed. On the right, Z represents some wire to which e^* moves the thinning link.



Handling the expansions is simpler, for one merely needs to prefix e^* (when necessary) with the rewiring steps which carry the thinning link across the expanded subgraph. For example, consider the following figure, in which a $(\otimes)$-expansion occurs on a wire with a thinning link.



In the above figure, Z represents some wire to which e^* moves the thinning link. e_x is the box-rewiring past the thinning link and the second occurrence of e is the same $\mathscr{E}$-move as the first, now possible since the link has been moved past the expansion. (It need not be necessary to move past the expansion, depending on where Z is located, of course.)

Finally we can use these skeletal translations to complete the proof, creating an $\mathscr{R}\cup\mathscr{X}\cup\mathscr{E}\cup\mathscr{E}^{\mathrm{op}}$-convergence for the $\mathscr{R}\cup\mathscr{X}$-divergence $n_1' \xrightarrow{v_1} n_1 \underset{\sim}{\xrightarrow{e^*}} n_2 \xrightarrow{v_2} n_2'$. We construct a "cube" as follows:

$$\begin{tikzcd}
 & n_1 \arrow[r, "e^*", "\sim"'] \arrow[dl, "v_1"'] \arrow[dr, dotted] & n_2 \arrow[dl, dotted] \arrow[dr, "v_2"] & \\
n'_1 \arrow[r, "{sk[v_1](e^*)}", "\sim"'] \arrow[dr, dotted] & n_{12} \arrow[dr, "v_2"] & n_{21} \arrow[r, "{sk[v_2](e^*)}", "\sim"'] \arrow[dl, "v_1"'] & n'_2 \arrow[dl, dotted] \\
 & n'_{21} \arrow[r, "\sim", "{sk[v_1;v_2](e^*)}"'] & n'_{12} &
\end{tikzcd}$$

The point here is that e^* might move some thinning links so as to prevent applying the reduction/expansion v_1 to n_2, and so we must free up the wires so we can apply v_1. The dotted arrows are reductions/expansions on skeletons, so involve no rewiring. The rewiring is done by the sk moves. The required confluence is obtained by following the solid arrows.

The orthogonality of $\mathscr{X}$ with $\mathscr{R}$ (that is, there are no overlapping rules) guarantees this is $\mathscr{X}$-reducing. □

Observe that a net can always be reduced to one whose skeleton is completely reduced. Two nets are then equivalent if (a) when so reduced they have the same skeleton and (b) one can use the rewiring equivalences to make their thinning link arrangements the same. Since there are only a finite number of possible configurations for the thinning links, this implies a search of all equivalent configurations is possible, providing a decision procedure.

Corollary 4.3. *There is a decision procedure for the equivalence of morphisms in* $\mathrm{Net}_{\emptyset}(\mathscr{C})$.

We should note that the $\otimes$–$\oplus$ expansion/reduction system, modulo the rest of the surgeries as $\mathscr{E}$-rules (including unit expansions and reductions), also yields an expansion-terminating expansion/reduction system modulo equations, in the non-commutative case as well as in the commutative case. That is:

Theorem 4.4 (Coherence: non-commutative case). *Let $\mathscr{C}$ be any polygraph; if we add Eqs.* (3)–(6) *and* (9)–(12) *to the equations (replace the arrows by equal signs), then for* $\mathrm{PNet}_\emptyset(\mathscr{C})$ *the system of tensor/cotensor reductions and expansions modulo this enlarged set of rewirings of thinning links is an expansion-terminating expansion/reduction system modulo equations (in the sense of Appendix* B).

Proof (sketch). This can be seen by looking closely at the proof of Theorem 4.1 above. This proof almost goes through unchanged for the non-symmetric case; the only place where we use symmetry is in assuming that we only need concern ourselves with the wires where thinning links are attached, and not the manner in which they are attached. This is only used in the unit reductions, where we essentially use the fact that in rewiring thinning links we only need keep in mind the preservation of the net criterion. This is not the case for the non-commutative case, where more care must be taken with the unit rewirings. Even if we assume empire moves (as in Proposition 3.3) we still cannot obtain confluence for unit reductions modulo the original $\mathscr{E}$. By simply adding the unit reductions and expansions to the equations, we can avoid this problem. □

Of course the addition of the unit reductions as equations also "avoids" having as a direct corollary a decision procedure for equality of morphisms in the non-symmetric case. We can no longer argue that the number of "rewirings" is finite, for one can add "unit bar-bells" (as in the proof of Proposition 3.1) ad libitum. To obtain a decision procedure requires a more detailed examination of these equivalences.

4.2. An example – the "triple-dual" problem

We can illustrate our techniques with a famous problem in coherence for monoidal closed categories, which in fact may be analyzed in the weakly distributive context. We start with the original diagram, in the monoidal closed context.

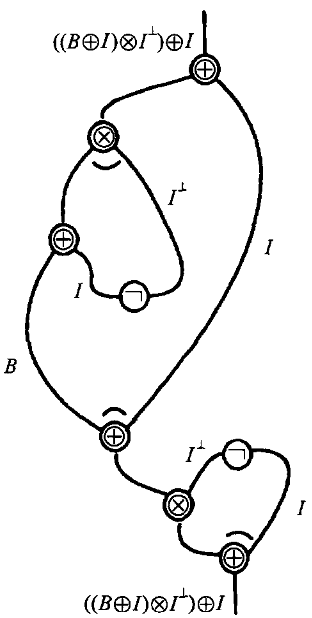
$$\begin{tikzcd}
((A \multimap I) \multimap I) \multimap I \arrow[r, "k_A \multimap id"] \arrow[dr, "id"'] & (A \multimap I) \arrow[d, "k_{A \multimap I}"] \\
 & ((A \multimap I) \multimap I) \multimap I
\end{tikzcd}$$

(where $k_A : A \longrightarrow ((A \multimap I) \multimap I)$ is the canonical such map, *viz.* the exponential transpose of evaluation).

In $*$-autonomous categories (or monoidal closed categories), this diagram generally does not commute. If I is not a unit, this is easily seen using the traditional treatment in terms of Kelly-Mac Lane graphs, but that method cannot be applied if I is a unit. One can verify that in fact this diagram will commute if $I = \bot$, generally does not commute if $I = \top$, but does commute if $A = I = \top$. (Note that without thinning links the expanded normal forms of the nets coincide in all cases when I is a unit, so that this is also an example that illustrates the need for thinning links, as well as the need to know how they may be rewired preserving equality of morphisms.)

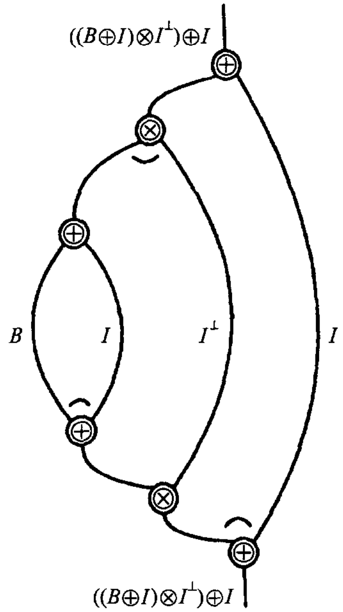
We note first that the particular instances of this diagram that we consider can be done in the weakly distributive context, where we replace I by a unit and $I^{\perp}$ with the other unit, and the negation links which would be necessary in the $*$-autonomous case (see Section 5) by the appropriate derived rules corresponding to the (iso)morphisms $\top \otimes \bot \longrightarrow \bot$ and $\top \longrightarrow \bot \oplus \top$. Of course, it is in the $*$-autonomous context that this diagram is historically interesting. It is in a sense the "simplest" diagram that cannot be handled by the traditional Kelly–Mac Lane techniques. In Section 5 we shall describe the negation links used in the diagram below, but the reader should keep in mind that we shall quickly pass to a version of this net that does not need such links.

We can define the internal hom as a derived operation: $A \multimap B = A^{\perp} \oplus B$. Then we can translate the composite $k_{A\multimap I} \circ k_A \multimap id$ above into a proof net; here is a step on the way to its expanded normal form (we write B for $A^{\perp}$ to prepare for the version of the net that may be constructed in the weakly distributive context).



(The negation ($\neg$) links are explained in Section 5; they can be replaced by appropriate subnets which make no use of any such new links in the cases where I is one unit and $I^{\perp}$ is the other unit.)

On the other hand, here is a similar step on the way to calculating the expanded normal form for the identity net.



In the nets above, if I were not a unit, these would be the expanded normal forms, and clearly these nets are not the same. An old idea of Lambek's [19] may be seen here: the "generality" of the first net is clearly a derivation of the sequent $((B \oplus C) \otimes C^{\perp}) \oplus D \longrightarrow ((B \oplus D) \otimes E^{\perp}) \oplus E$, whereas the generality of the second is $((B \oplus C) \otimes D) \oplus E \longrightarrow ((B \oplus C) \otimes D) \oplus E$. This is no surprise; it is exactly what one would expect if I was not a unit. Next consider the case if I is a unit.

If I is the unit $\top$, say, then the nodes at I and $I^{\perp}$ have to be expanded, since in expanded normal form, each occurrence of a unit (recall $\top^{\perp} = \perp$) must either come from or go to a null node. This in effect transforms several of the edges in the graphs above into thinning links.

We leave it as an exercise to show that in the case when $I = \top$ (but A arbitrary) each discharged unit has a trivial (singleton) empire, and so no rewiring is possible; hence the diagram does not commute. And similarly that the rewiring may be done if $I = \perp$, so that the diagram does commute. But now consider the case where $A = I = \top$ (where the diagram commutes). We must show how to rewire the net corresponding to the compound morphism to give the identity. This is shown in Fig. 3: the point here is that with $A = \top$ (corresponding to the wire/thinning link for $\perp$ at the left) we have an extra discharged unit, whose empire is not trivial. Although it would seem that we do not have to move this thinning link, by doing so we make other rewirings possible, since the empires of other units may change with each rewiring. Once the required rewirings are done, the initial rewiring is reversed to finish with the expanded normal form of the identity map.

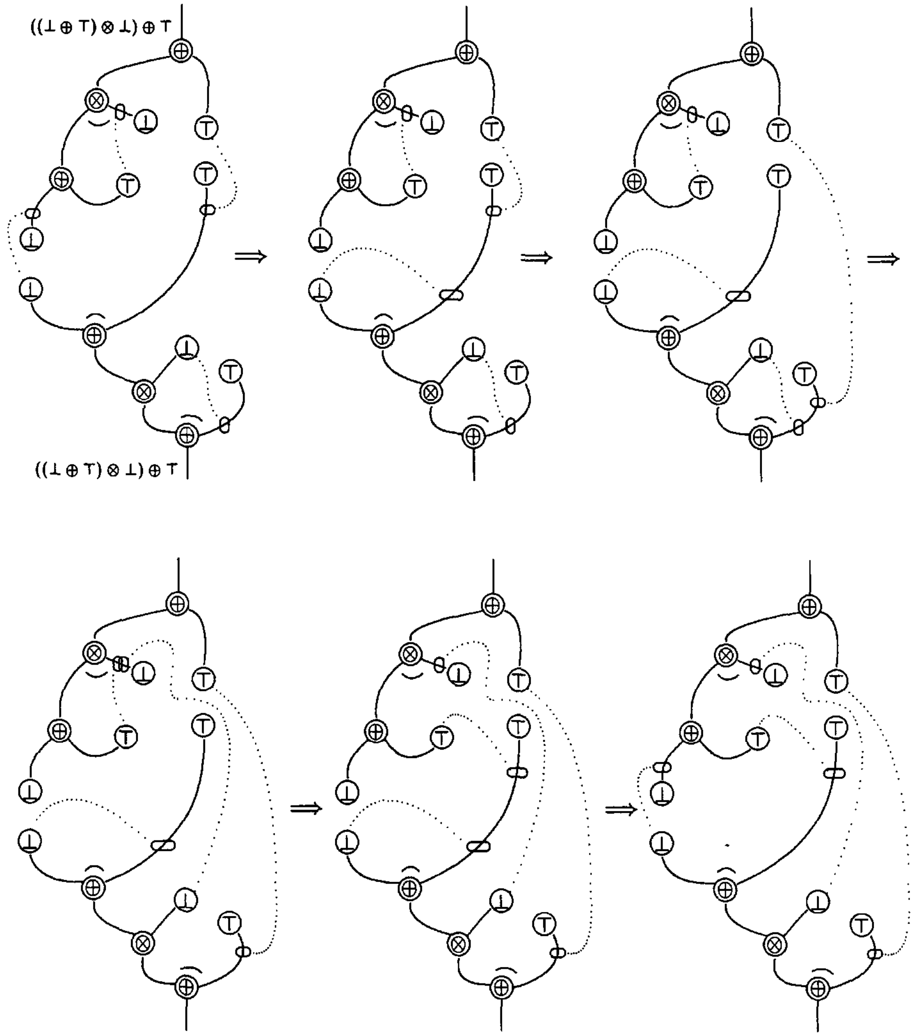


Fig. 3. An example of rewiring.

A further exercise for the reader would be to identify the basic rewirings (in the sense of Section 3) necessary to directly establish the equivalence of these nets, without using the empire criterion. In fact, this example (suitably modified) works in the non-commutative case with the rewiring steps from Section 3.

5. Adding negation

We shall now illustrate how to extend our nets to include negation. The key ideas needed are already available as we can just add, as was implicit in [11], the appropriate components and net equivalences. The new components correspond to the non-logical axioms

$$A \otimes A^{\perp} \longrightarrow \perp \qquad\qquad \top \longrightarrow A^{\perp} \oplus A$$

which together with the required equivalences were described in Definition 1.2. The developments of this paper allow us to make two significantly new observations. The first follows from the fact that the expansion/reduction system of Theorems 4.1, 4.4 can be extended to the $*$-autonomous case; this provides a decision procedure for map equality in free symmetric $*$-autonomous categories. The second is that the free $*$-autonomous category built from a weakly distributive category by adding negation contains as a full and faithful subcategory the weakly distributive category from which it is built. This is of some interest as weakly distributive categories include a rather wide variety of categories (from commutative rings to braided monoidal categories with shift objects) implying that all these can be fully and faithfully "completed" to $*$-autonomous categories.

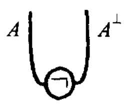
5.1. Generating the free $*$-autonomous category

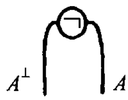
To illustrate the method, we shall construct the free symmetric $*$-autonomous category generated by a polygraph with negation. A similar treatment can be given for the non-symmetric case: it is more complex as two different negations must be accommodated (see [6]) with accompanying coherence diagrams.

We may start as in Section 2.2, with a set of atomic types $\mathscr{A}$, but this time we generate all the linear types. This necessitates adding the rule

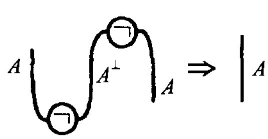
- If A is a formula then $A^\perp$ is a formula.

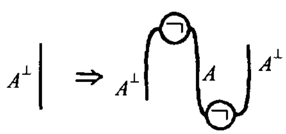
From these types we may then generate nets as before. Negation is obtained by adding for each formula components for negation, accompanied by net equivalences (we present the equivalences as rewrites for later reference – the reader can substitute equal signs as necessary):

(γ) 

(τ) 

Here are the corresponding equivalences (rewrites):

($\neg$- reduction) 

($\neg$- expansion) 

As rewrites in our term notation these become (where τ and γ refer to the components as above):

$$A : \langle x||x\rangle : A \Leftarrow A : \langle x|[\,]\tau[z,y];[x,z]\gamma[\,]|y\rangle : A \tag{13}$$

$$A^\perp : \langle x||x\rangle : A^\perp \Rightarrow A^\perp : \langle x|[\,]\tau[x,z];[z,x]\gamma[\,]|x\rangle : A^\perp \tag{14}$$

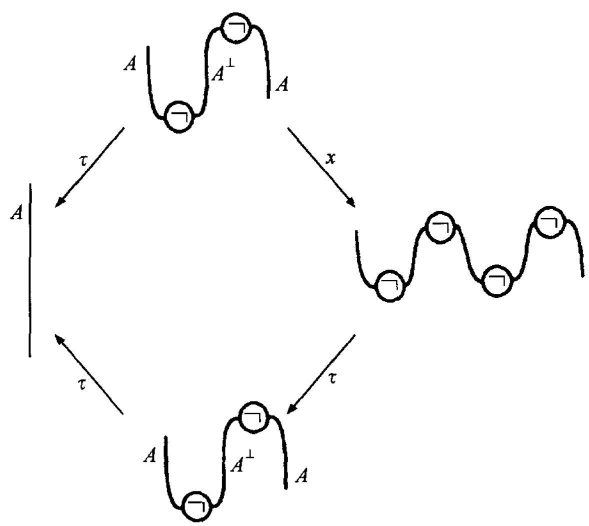
We shall call the category which results from this construction (with additional components $\mathscr{C}$ and equivalences E) $\mathrm{Net}^*_E(\mathscr{C})$. In view of Theorem 1.3 it is now immediate that:

Theorem 5.1. $\mathrm{Net}^*_E(\mathscr{C})$ *is the free* $*$*-autonomous category generated by the polygraph* (*with negation*) $\mathscr{C}$ *and the equivalences* E.

Furthermore, treating the equivalences introduced for $*$-autonomy above as reduction and expansion rewrites respectively, we obtain:

Theorem 5.2. *The system of reductions* (1–6, 13) *and expansions* (7–12, 14) *is a strongly normalizing expansion/reduction system modulo the rewiring equations* (15–60) *on* $\mathrm{Net}^*_\emptyset(\mathscr{C})$.

Proof (sketch). On the skeleton we still have a strongly normalizing expansion-reduction system. The new rules are cut-conservative and only interact with each other to produce the following critical pair:



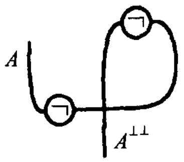
The introduced $\neg$-expansion is clearly promoting and so the system is expansion-terminating (a double expansion will precipitate a reduction).

As negations are components we may freely rewire over them implying that the extensions to skeleton maps (as in the proof of Theorem 4.1) can be carried through. $\square$

As the residual equivalence is determined by rewirings whose equivalence classes are finite, we can now conclude in the symmetric case:

Corollary 5.3. *There is a decision procedure for the equality of the morphisms of free* $*$*-autonomous categories*, $\mathrm{Net}^*_\emptyset(\mathscr{C})$.

Perhaps we should draw attention to one point about our presentation. We have presented a natural deduction system for $*$-autonomous categories, i.e. for mLL including the units. Of course, we can simulate one-sided nets, by having only empty nodes as source (or initial) nodes, and indeed, in view of the adjunctions of [6], we can translate a two-sided net into such a simulation by moving initial nodes to the corresponding negated terminal nodes. Essentially this corresponds to "capping" the net with a $(\otimes E)$ link and a (τ) link. But our two-sided nets for mLL are not quite the same as Girard's one-sided nets, even via the simulation, since, for example, with the two-sided nets one has an isomorphism $A \xrightarrow{\sim} A^{\perp\perp}$ rather than an identity. Girard introduced negation via a series of such abbreviations, which we believe ought to be seen as inessential. Certainly the general situation in $*$-autonomous categories is that such maps are *not* identities. So the two-sided nets reflect the reality of $*$-autonomous categories more accurately. The extra τ and γ links would allow similar nets in the planar case, but then the isomorphisms created would be between A and either ${}^{\perp}(A^{\perp})$ or $({}^{\perp}A)^{\perp}$. For the record, here is the isomorphism referred to.



5.2. A conservative extension

As the condition of sequentiality is the same for $*$-autonomous and weakly distributive categories it follows that the rewiring equivalence class of a skeletal normal form weakly distributive morphism is the same whether it sits in the $*$-autonomous setting or the weakly distributive setting. This shows that the embedding of a free weakly distributive category into its free $*$-autonomous completion is faithful. Furthermore, the ability to eliminate all but the "essential cuts" implies that this embedding is full.

Logically, this implies that the addition of negation to the linear sequent calculus is a conservative extension (in the sense that no new theorems are added). However, it is a very much stronger proof-theoretic statement as it also addresses the finer structure of proof equivalence: that is whether two deductions can become equivalent under the extension (faithful), or indeed whether a completely new proof becomes available (full).

From the categorical perspective, however, this begs a larger question: can an arbitrary (non-free) weakly distributive category always be completed in such a full and faithful manner to a $*$-autonomous category? This can be stated as the requirement that the unit of the adjunction, between weakly distributive and $*$-autonomous categories, is full and faithful.

First, define **WDC** to be the category of weakly distributive categories and weakly distributive functors. Next, let $*$-**AUT** be the category of $*$-autonomous categories and $*$-autonomous functors. There is an evident forgetful functor from $*$-**AUT** to **WDC**. By formally adjoining negation to a weakly distributive category, we obtain a left adjoint to this forgetful functor. We now state our conservativity result as follows.

Theorem 5.4. *Let* $\mathbf{C}$ *be a weakly distributive category. If* F *is the above-mentioned free functor, and* η *is the unit of the adjunction,*

$$\eta : \mathbf{C} \longrightarrow U(F(\mathbf{C})),$$

then η *is full and faithful.*

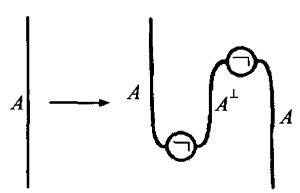
Proof (*fullness*). Given a morphism $f{:}X \longrightarrow Y$ in $U(F(\mathbf{C}))$, where $X = \eta(X')$, $Y = \eta(Y')$ for X', Y' in C, note first that X', Y' and hence X, Y do not involve any negation operator. So if one considers any normalized net presentation $\mathcal{N}$ of f, it must have no instances of negation links (τ), (γ) (because it is in normal form). Since the net criterion is the same in the two theories, $\mathcal{N}$ must be a valid net presentation of a morphism $f'{:}X' \longrightarrow Y'$ so that $\eta(f') = f$.

(*Faithfulness*). Suppose we have an equivalence of morphisms in $U(F(\mathbf{C}))$:

$$f = g{:}X \longrightarrow Y$$

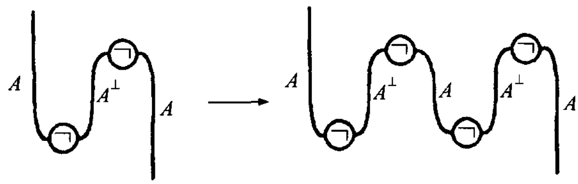
and that f and g are in the image of $\mathbf{C}$. We wish to show that f and g are equivalent in $\mathbf{C}$. (Note that we are ignoring the distinction between a map and its image under η.) Let $\mathcal{N}_f$ and $\mathcal{N}_g$ denote weakly distributive nets corresponding to f and g, respectively. Note in particular that there are no negation links or negated formulas in $\mathcal{N}_f$ or $\mathcal{N}_g$. Since it is the case that $f = g$ in $U(F(\mathbf{C}))$, there must be a sequence of net rewrites from $\mathcal{N}_f$ to $\mathcal{N}_g$ in $U(F(\mathbf{C}))$. These rewrites may involve the negation links and the γ and τ rewrites and their inverses. We need to show that there is a corresponding sequence of rewrites in $\mathbf{C}$, i.e. we need a sequence of rewrites not involving τ or γ.

So, consider a sequence of rewrites from $\mathcal{N}_f$ to $\mathcal{N}_g$, possibly involving γ and τ. The idea will be to rearrange the sequence of rewrites so that the γ and τ rewrites all occur at the end of the sequence, and thus cancel each other out. Since the initial net $\mathcal{N}_f$ does not contain any negation, the first instance of a negation rewrite must be of the following form (an inverse ($\neg$)-reduction):

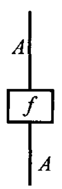


We wish to determine what sort of rewrites can interact with the negated parts of the structure. The rewrites which are relevant to our sequence are of the following form:

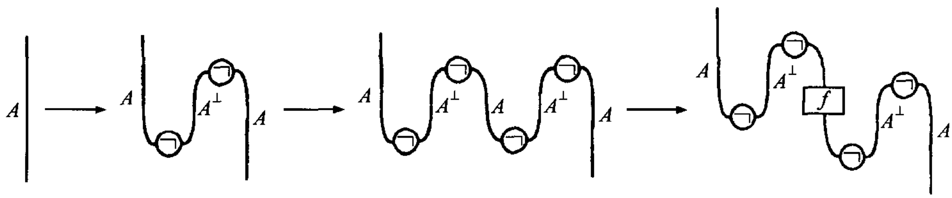
1. The A "wire" can be replaced by any appropriate subgraph according to a rewrite valid in C, or
2. The $A^\perp$ wire can be expanded via a negation link.



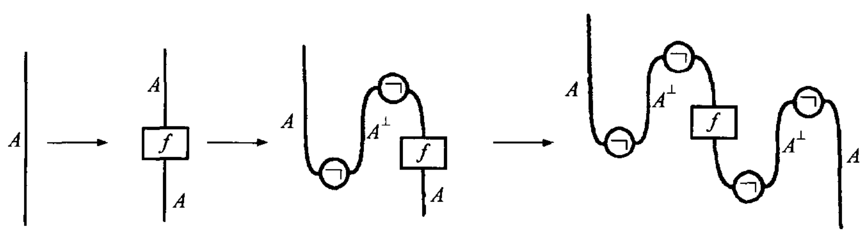
Suppose the next rewrite acts on the A wire, replacing A with another subnet representing an endomorphism on A, say,



where the square f represents an arbitrary subnet. The technique for the case when an arbitrary subnet representing a morphism $A \longrightarrow B$ in **C** is replaced with another (equivalent) subnet representing the same morphism is similar. Then we would replace a subsequence of rewrites of this form



with a subsequence of this form:



In a similar manner, we can "push" a $*$-rewrite past any other **C**-rewrite. (The other forms to check are those involving the tensors or units – and these do not overlap the negation rewrites – and those involving commutative diagrams of **C**, which are treated as atomic rewrites and behave just as the rewrite on the identity A above.) In this manner, by induction one can show that we can replace the original sequence with a sequence such that all of the γ and τ rewrites occur at the end of the sequence.

So, the sequence is now of the following form:

$$\mathcal{N}_f = \mathcal{N}_{f_0} \Rightarrow \mathcal{N}_{f_1} \Rightarrow \cdots \Rightarrow \mathcal{N}_{f_k} \Rightarrow \mathcal{N}_{g_1} \Rightarrow \cdots \Rightarrow \mathcal{N}_{g_n} = \mathcal{N}_g$$

where the rewrites $\mathcal{N}_{f_i} \Rightarrow \mathcal{N}_{f_{i+1}}$ are "negation-free". Since the net $\mathcal{N}_g$ is a weakly distributive net, it must be the case that all of the rewrites at the end involving negation links cancel; indeed, the net $\mathcal{N}_{f_k}$ must be trivially equivalent to $\mathcal{N}_g$, and so represents the morphism g. So this is the sequence establishing the equivalence of f and g in **C**. $\square$

Appendix A. Surgery for thinning and cothinning links

These are the rules concerned with the manipulation of the thinning and cothinning links. We divide them into those for the interaction of thinning and the tensor and cotensor links, cothinning and the tensor and cotensor links, thinning and cothinning, and, finally, thinning, cothinning and an arbitrary component.

Recall that there is an implicit assumption in these rules: *a rule of surgery can only be applied to a proof net if it preserves the net criterion; that is, if after the surgical alteration, one still has a proof net.* Therefore, there is a hidden cost in applying these rules: namely one must check that the surgical alteration yields a sequential net when rewiring past a switching link. Explicitly, Eqs. (16), (17), (22), (23), (25), (26), (31), and (32) are the surgery rules that require one to check that the right-hand side is a net if the left-hand side is.

The rules for the unit start with the obvious ones for manipulations over the tensor, followed by less obvious rules for the interaction (as demanded by weak distributivity) of the unit with the cotensor:

$$\begin{aligned} &A,\top,B : \langle x,z,y|[x,z]\top E^R[x];[x,y]\otimes I[w]|w\rangle : A\otimes B \\ &\quad = A,\top,B : \langle x,z,y|[z,y]\top E^L[y];[x,y]\otimes I[w]|w\rangle : A\otimes B \end{aligned} \tag{15}$$

$$\begin{aligned} &\top,A\otimes B : \langle z,x|[z,x]\top E^L[x];[x]\otimes E[x_1,x_2]|x_1,x_2\rangle : A,B \\ &\quad = \top,A\otimes B : \langle z,x|[x]\otimes E[x_1,x_2];[z,x_1]\top E^L[x_1]|x_1,x_2\rangle : A,B \end{aligned} \tag{16}$$

$$\begin{aligned} &A\otimes B,\top : \langle x,z|[x,z]\top E^R[x];[x]\otimes E[x_1,x_2]|x_1,x_2\rangle : A,B \\ &\quad = A\otimes B,\top : \langle x,z|[x]\otimes E[x_1,x_2];[x_2,z]\top E^R[x_2]|x_1,x_2\rangle : A,B \end{aligned} \tag{17}$$

$$A,B,\top : \langle x_1,x_2,z|[x_2,z]\top E^R[x_2];[x_1,x_2]\otimes I[x]|x\rangle : A\otimes B$$
$$= A,B,\top : \langle x_1,x_2,z|[x_1,x_2]\otimes I[x];[x,z]\top E^R[x]|x\rangle : A\otimes B \tag{18}$$
$$\top,A,B : \langle z,x_1,x_2|[z,x_1]\top E^L[x_1];[x_1,x_2]\otimes I[x]|x\rangle : A\otimes B$$
$$= \top,A,B : \langle z,x_1,x_2|[x_1,x_2]\otimes I[x];[z,x]\top E^L[x]|x\rangle : A\otimes B \tag{19}$$
$$\top,A\oplus B : \langle z,x|[z,x]\top E^L[x];[x]\oplus E[x_1,x_2]|x_1,x_2\rangle : A,B$$
$$= \top,A\oplus B : \langle z,x|[x]\oplus E[x_1,x_2];[z,x_1]\top E^L[x_1]|x_1,x_2\rangle : A,B \tag{20}$$
$$A\oplus B,\top : \langle x,z|[x,z]\top E^R[x];[x]\oplus E[x_1,x_2]|x_1,x_2\rangle : A,B$$
$$= A\oplus B,\top : \langle x,z|[x]\oplus E[x_1,x_2];[x_2,z]\top E^R[x_2]|x_1,x_2\rangle : A,B \tag{21}$$
$$A,B,\top : \langle x_1,x_2,z|[x_2,z]\top E^R[x_2];[x_1,x_2]\oplus I[x]|x\rangle : A\oplus B$$
$$= A,B,\top : \langle x_1,x_2,z|[x_1,x_2]\oplus I[x];[x,z]\top E^R[x]|x\rangle : A\oplus B \tag{22}$$
$$\top,A,B : \langle z,x_1,x_2|[z,x_1]\top E^L[x_1];[x_1,x_2]\oplus I[x]|x\rangle : A\oplus B$$
$$= \top,A,B : \langle z,x_1,x_2|[x_1,x_2]\oplus I[x];[z,x]\top E^L[x]|x\rangle : A\oplus B \tag{23}$$

The rules for counit cothinning are dual, we list them for completeness. We start, as before, with the obvious ones for manipulations over the cotensor links, followed by less obvious rules for the interaction (as demanded by weak distributivity) of the cothinning link with the tensor links:

$$A\oplus B : \langle w|[w]\oplus E[x,y];[x]\bot E^R[x,z];|x,z,y\rangle : A,\bot,B$$
$$= A\oplus B : \langle w|[w]\oplus E[x,y];[y]\bot E^L[z,y];|x,z,y\rangle : A,\bot,B \tag{24}$$
$$A,B : \langle x_1,x_2|[x_1,x_2]\oplus I[x];[x]\bot I^R[x,z]|x,z\rangle : A\oplus B,\bot$$
$$= A,B : \langle x_1,x_2|[x_2]\bot I^R[x_2,z];[x_1,x_2]\oplus I[x]|x,z\rangle : A\oplus B,\bot \tag{25}$$
$$A,B : \langle x_1,x_2|[x_1,x_2]\oplus I[x];[x]\bot I^L[z,x]|x,z\rangle : \bot,A\oplus B$$
$$= A,B : \langle x_1,x_2|[x_1]\bot I^L[z,x_1];[x_1,x_2]\oplus I[x]|x,z\rangle : \bot,A\oplus B \tag{26}$$
$$A\oplus B : \langle x|[x]\oplus E[x_1,x_2];[x_2]\bot I^R[x_2,z]|x_1,x_2,z\rangle : A,B,\bot$$
$$= A\oplus B : \langle x|[x]\bot I^R[x,z];[x]\oplus E[x_1,x_2]|x_1,x_2,z\rangle : A,B,\bot \tag{27}$$
$$A\oplus B : \langle x|[x]\oplus E[x_1,x_2];[x_1]\bot I^L[z,x_1]|z,x_1,x_2\rangle : \bot,A,B$$
$$= A\oplus B : \langle x|[x]\bot I^R[z,x];[x]\oplus E[x_1,x_2]|z,x_1,x_2\rangle : \bot,A,B \tag{28}$$
$$A,B : \langle x_1,x_2|[x_1,x_2]\otimes I[x];[x]\bot I^R[x,z]|x,z\rangle : A\otimes B,\bot$$
$$= A,B : \langle x_1,x_2|[x_2]\bot I^R[x_2,z];[x_1,x_2]\otimes I[x]|x,z\rangle : A\otimes B,\bot \tag{29}$$

$$A,B : \langle x_1,x_2|[x_1,x_2] \otimes I[x]; [x]\bot I^L[z,x]|x,z\rangle : \bot, A \otimes B$$
$$= A,B : \langle x_1,x_2|[x_1]\bot I^L[z,x_1]; [x_1,x_2] \otimes I[x]|x,z\rangle : \bot, A \otimes B \tag{30}$$
$$A \otimes B : \langle x|[x] \otimes E[x_1,x_2]; [x_2]\bot I^R[x_2,z]|x_1,x_2,z\rangle : A,B,\bot$$
$$= A \otimes B : \langle x|[x]\bot I^R[x,z]; [x] \otimes E[x_1,x_2]|x_1,x_2,z\rangle : A,B,\bot \tag{31}$$
$$A \otimes B : \langle x|[x] \otimes E[x_1,x_2]; [x_1]\bot I^L[z,x_1]|z,x_1,x_2\rangle : \bot,A,B$$
$$= A \otimes B : \langle x|[x]\bot I^L[z,x]; [x] \otimes E[x_1,x_2]|z,x_1,x_2\rangle : \bot,A,B \tag{32}$$

The rules governing the interaction of the unit thinning and counit cothinning are as follows:

$$\top,A : \langle z,x|[z,x]\top E^L[x]; [x]\bot I^R[x,z']|x,z'\rangle : A,\bot$$
$$= \top,A : \langle z,x|[x]\bot I^R[x,z']; [z,x]\top E^L[x]|x,z'\rangle : A,\bot \tag{33}$$
$$A,\top : \langle x,z|[x,z]\top E^R[x]; [x]\bot I^L[z',x]|x,z'\rangle : \bot,A$$
$$= A,\top : \langle z,x|[x]\bot I^R[x,z']; [z,x]\top E^L[x]|z',x\rangle : \bot,A \tag{34}$$
$$\top,A : \langle z,x|[z,x]\top E^L[x]; [x]\bot I^L[z',x]|z',x\rangle : \bot,A$$
$$= \top,A : \langle z,x|[z]\bot I^L[z',z]; [z,x]\top E^L[x]|z',x\rangle : \bot,A \tag{35}$$
$$= \top,A : \langle z,x|[x]\bot I^L[z',x]; [z,z']\top E^L[z']|z',x\rangle : \bot,A \tag{36}$$
$$A,\top : \langle x,z|[x,z]\top E^R[x]; [x]\bot I^R[x,z']|x,z'\rangle : A,\bot$$
$$= A,\top : \langle x,z|[z]\bot I^R[z,z']; [x,z]\top E^R[x]|x,z'\rangle : A,\bot \tag{37}$$
$$= A,\top : \langle x,z|[x]\bot I^R[x,z']; [z',z]\top E^R[z']|z',x\rangle : A,\bot \tag{38}$$

The interaction of thinning and thinning, and cothinning and cothinning are as follows:

$$\top,\top,A : \langle z_1,z_2,x||x\rangle : A$$
$$= \top,\top,A : \langle z_1,z_2,x|[z_1,z_2]\top E^L[z_2]; [z_2,x]\top E^L[x]|x\rangle : A \tag{39}$$
$$= \top,\top,A : \langle z_1,z_2,x|[z_1,z_2]\top E^R[z_1]; [z_1,x]\top E^L[x]|x\rangle : A \tag{40}$$
$$A,\top,\top : \langle x,z_1,z_2|[x,z_1]\top E^R[x]; [x,z_2]\top E^R[x]|y\rangle : A$$
$$= A,\top,\top : \langle x,z_1,z_2|[z_1,z_2]\top E^R[z_1]; [x,z_1]\top E^R[x]|x\rangle : A \tag{41}$$
$$= A,\top,\top : \langle x,z_1,z_2|x,z_1,z_2|[z_1,z_2]\top E^L[z_2]; [x,z_2]\top E^R[x]|x\rangle : A \tag{42}$$
$$\top,A,\top : \langle z_1,x,z_2|[z_1,x]\top E^L[x]; [x,z_2]\top E^R[x]|y\rangle : A$$
$$= \top,A,\top : \langle z_1,x,z_2|[x,z_2]\top E^R[x]; [z_1,x]\top E^L[x]|x\rangle : A \tag{43}$$

$$A : \langle x|[x]\bot E^L[z_2,x];[x]\bot E^L[z_1,x]|z_1,z_2,x\rangle : \bot,\bot,A$$
$$= A : \langle x|[x]\bot E^L[z_2,x];[z_2]\bot E^L[z_1,z_2]|z_1,z_2,x\rangle : \bot,\bot,A \tag{44}$$
$$= A : \langle x|[x]\bot E^L[x,z_1];[z_1]\bot E^R[z_1,z_2]|z_1,z_2,y\rangle : \bot,\bot,A \tag{45}$$
$$A : \langle x|[x]\bot E^R[x,z_2];[x]\bot E^R[x,z_1]|x,z_1,z_2\rangle : A,\bot,\bot$$
$$= A : \langle x|[x]\bot E^R[x,z_1];[z_1]\bot E^R[z_1,z_2]|x,z_1,z_2\rangle : A,\bot,\bot \tag{46}$$
$$= A : \langle x|[x]\bot E^R[x,z_2];[z_2]\bot E^L[z_1,z_2]|x,z_1,z_2\rangle : A,\bot,\bot \tag{47}$$
$$A : \langle x|[x]\bot E^R[x,z_2];[x]\bot E^L[z_1,x]|z_1,x,z_2\rangle : \bot,A,\bot$$
$$= A : \langle x|[x]\bot E^L[z_1,x];[x]\bot E^R[x,z_2]|z_1,x,z_2\rangle : \bot,A,\bot \tag{48}$$

Finally the rules governing interaction of the units with a component (we call these the "box-rewiring" rules) are as follows, taking the unit first:

$$\Gamma_1,A,\top,B,\Gamma_2 : \langle ..,x_1,z,x_2,..|[x_1,z]\top E^R[x_1];[..,x_1,x_2,..]f[..]|..\rangle : \Delta$$
$$= \Gamma_1,A,\top,B,\Gamma_2 : \langle ..,x_1,z,x_2,..|[z,x_2]\top E^L[x_2];[..,x_1,x_2,..]f[..]|..\rangle : \Delta \tag{49}$$
$$\top,A,\Gamma : \langle z,x_1,..|[z,x_1]\top E^L[x_1];[x_1,..]f[x_2,..]|x_2,..\rangle : B,\Delta$$
$$= \top,A,\Gamma : \langle z,x_1,..|[x_1,..]f[x_2,..];[z,x_2]\top E^L[x_2]|x_2,..\rangle : B,\Delta \tag{50}$$
$$\Gamma,A,\top : \langle ..,x_1,z|[x_1,z]\top E^R[x_1];[..,x_1]f[..,x_2]|..,x_2\rangle : \Delta,B$$
$$= \Gamma,A,\top : \langle ..,x_1,z|[..,x_1]f[..,x_2];[x_2,z]\top E^R[x_2]|..,x_2\rangle : \Delta,B \tag{51}$$
$$\top : \langle z|[\,]f[x_1,..,x_2];[z,x_1]\top E^L[x_1]|x_1,..,x_2\rangle : A,\Delta,B$$
$$= \top : \langle z|[\,]f[x_1,..,x_2];[x_2,z]\top E^R[x_2]|x_1,..,x_2\rangle : A,\Delta,B \tag{52}$$

and for the counit:

$$\Gamma : \langle ..|[..]f[..,x_1,x_2,..];[x_1]\bot E^R[x_1,z]|..,x_1,z,x_2,..\rangle : \Delta_1,A,\bot,B,\Delta_2$$
$$= \Gamma : \langle ..|[..]f[..,x_1,x_2,..];[x_2]\bot E^L[z,x_2]|..,x_1,z,x_2,..\rangle : \Delta_1,A,\bot,B,\Delta_2 \tag{53}$$
$$A,\Gamma : \langle x_1,..|[x_1,..]f[x_2,..];[x_2]\bot E^L[z,x_2]|z,x_2,..\rangle : \bot,B,\Delta$$
$$= A,\Gamma : \langle x_1,..|[x_1]\bot E^L[z,x_1];[x_1,..]f[x_2,..]|z,x_2,..\rangle : \bot,B,\Delta \tag{54}$$
$$\Gamma,A : \langle x_1,..|[x_1,..]f[x_2,..];[x_2]\bot E^R[x_2,z]|..,x_2,z\rangle : \Delta,B,\bot$$
$$= \Gamma,A : \langle ..,x_1|[x_1]\bot E^R[x_1,z];[x_1,..]f[x_2,..]|..,x_2,z\rangle : \Delta,B,\bot \tag{55}$$
$$A,\Gamma,B : \langle x_1,..,x_2|[x_1]\bot I^L[z,x_1];[x_1,..,x_2]f[\,]|z\rangle : \bot$$
$$= A,\Gamma,B : \langle x_1,..,x_2|[x_2]\bot I^R[x_2,z];[x_1,..,x_2]f[\,]|z\rangle : \bot \tag{56}$$

To handle non-planar nets all we need do is add four extra rules concerning the units. The rest is handled by the change in the circuits (to allow wire crossings) and the definition of sequentiality. The first rules of surgery concern the new relationship between the unit and counit eliminations.

$$\top,A : \langle z,x|[z,x]\top E^L[x]|x\rangle : A = \top,A : \langle z,x|[x,z]\top E^R[x]|x\rangle : A \tag{57}$$

$$A : \langle x|[x]\bot I^L[z,x]|z,x\rangle : \bot,A = A : \langle x|[x]\bot I^R[x,z]|z,x\rangle : \bot,A \tag{58}$$

Notice that these two identities already use the ability to cross wires. The next two rules concern the interaction of the units with the arbitrary components. Notice that a unit elimination can be moved down the sides of a component and from one wire to the next along the top of a component – here we must use the introduced rule. The rules below allow us to move the unit elimination across the bottom and dually the counit elimination along the top.

$$\begin{aligned}&\Gamma,\top : \langle ..,z|[..]f[..,x_1,x_2,..];[x_1,z]\top E^R[x_1]|..,x_1,x_2,..\rangle : \Delta_1,A,B,\Delta_2\\&\quad = \Gamma,\top : \langle ..|[..]f[..,x_1,x_2,..];[x_2,z]\top E^L[x_2]|..,x_1,x_2,..\rangle : \Delta_1,A,B,\Delta_2 \qquad (59)\\&\Gamma_1,A,B,\Gamma_2 : \langle ..,x_1,x_2,..|[x_1]\bot E^R[x_1,z];[..,x_1,x_2,..]f[..]|z,..\rangle : \bot,\Delta\\&\quad = \Gamma_1,A,B,\Gamma_2 : \langle ..,x_1,x_2,..|[x_2]\bot E^L[z,x_2];[..,x_1,x_2,..]f[..]|z,..\rangle : \bot,\Delta \qquad (60)\end{aligned}$$

Thus a thinning link, in the symmetric case, attached to a component can be moved to any wire leaving (or entering) that component. This is in stark contrast to the somewhat restricted movement permitted in the planar case.

Recall that a representative sample of these thinning link moves (or rewirings) is displayed graphically in Fig. 2.

Appendix B. Notes on expansion/reduction systems modulo equations[7]

The results in this paper rely heavily on the use of expansion/reduction rewriting in the presence of equations. The purpose of this section is to give an exposition of the general theory of these rewrite systems and to introduce their terminology. We divide the section into three parts. Section B.1 introduces the basic terminology of rewriting. Section B.2 discusses expansion/reduction rewrite systems in the absence of equations, and Section B.3 discusses the modifications needed when equations are added.

B.1. Rewrite systems

A *rewrite system*, $\mathscr{R} = (N,R,\partial_0^R,\partial_1^R)$, on a set N consists of a directed graph[8] with nodes N and arrows R, each of which has a domain and codomain, given by respectively ∂_0^R and ∂_1^R. An arrow $n_1 \xrightarrow{r} n_2$ is to be thought of as a reduction. We shall say that the rewrite system is *terminating* at $n \in N$ in case there is a (least) $\mathrm{bnd}_{\mathscr{R}}(n) \in \mathbb{N}$ which bounds the length of any rewriting chain $n \underset{*}{\longrightarrow} n'$. A rewrite system is *terminating* when it terminates at every $n \in N$.

[7] By Robin Cockett.

[8] Rewriting systems are usually regarded to be relations between the terms. By allowing directed graphs one allows the possibility of multiple ways of getting from one node to another which is a useful generalization.

An $n \in N$ is a *$\mathscr{R}$-normal form* when $\mathrm{bnd}_{\mathscr{R}}(n)$ exists and is zero. When one is using a rewriting system to provide a decision procedure for the equivalence relation generated by the image of the arrows, $\mathscr{R}^{=}$, the objective is to arrange that there is exactly one normal form in each equivalence class. The key notion in achieving this is confluence: a rewriting system is *confluent* if for each *divergence*

$$n_1 \underset{*}{\longleftarrow} n \underset{*}{\longrightarrow} n_2$$

there is a *convergence*

$$n_1 \underset{*}{\longrightarrow} n' \underset{*}{\longleftarrow} n_2.$$

It is a classic result of rewriting theory that when $\mathscr{R}$ is terminating one need only check *local confluence*, that is that every single step divergence

$$n_1 \longleftarrow n \longrightarrow n_2$$

has a convergence of the above form. Furthermore, when $\mathscr{R}$ is specified by redex-reduct pairs, then it is sufficient to check this local condition only for *critical divergences*. When the system is specified by finitely many redex–reduct pairs, and where the only critical divergences arise from overlapping redexes, since then there are then only finitely many local critical divergences, the determination of local confluence becomes feasible.

A rewrite system $\mathscr{R}$ which is confluent has at most one normal form in each $\mathscr{R}^{=}$ equivalence class. A rewriting system $\mathscr{R}$ is said to be *normalizing at $n \in N$* if there is a unique normal form in the equivalence class of n, $\mathrm{nf}_{\mathscr{R}}(n)$, and it can be reached from n by a rewrite chain

$$n \underset{*}{\longrightarrow} \mathrm{nf}_{\mathscr{R}}(n).$$

More generally a rewrite system $\mathscr{R}$ is said to be *normalizing* if it is normalizing at each $n \in N$. A normalizing rewriting system is necessarily confluent.

A terminating confluent rewriting system is certainly normalizing but has a further crucial property: namely that to reach the normal form of n one can rewrite in any order. In a general normalizing system this may not be the case. We shall call a terminating confluent rewriting system a *reduction system* and refer to the normal forms of such systems as *reduced normal forms*.

B.2. Expansion/reduction rewrite systems

An *expansion/reduction system $\mathscr{X}/\mathscr{R}$* on N consists of a reduction system $\mathscr{R}$ on N together with a rewrite system $\mathscr{X}$ on N such that the joint system $\mathscr{X} \cup \mathscr{R}$ is confluent and *$\mathscr{X}$-reducing*, that is any divergence

$$n_1 \overset{v_1}{\underset{*}{\longleftarrow}} n \overset{v_2}{\underset{*}{\longrightarrow}} n_2$$

has a convergence

$$n_1 \xrightarrow[*]{v_2'} n' \xleftarrow[*]{v_1'} n_2$$

such that $\mathrm{len}_{\mathscr{X}}(v_i) \geq \mathrm{len}_{\mathscr{X}}(v_i')$ for $i = 1, 2$ (where $\mathrm{len}_{\mathscr{X}}(v)$ is the number of $\mathscr{X}$ steps in v).

An expansion/reduction system, $\mathscr{X}/\mathscr{R}$ is said to be *expansion terminating at* (or $\mathscr{X}/\mathscr{R}$ terminating at) $n \in N$ in case there is a (least) bound $\mathrm{bnd}_{\mathscr{X}/\mathscr{R}}(n)$ on the number of irreducible expansions, $\mathrm{len}_{\mathscr{X}/\mathscr{R}}(v)$, of any chain v of $\mathscr{X} \cup \mathscr{R}$ rewrites leaving n. An *irreducible expansion* is a one-step expansion $n \xrightarrow{x} n'$ with $\mathrm{nf}_{\mathscr{R}}(n) \neq \mathrm{nf}_{\mathscr{R}}(n')$. The whole expansion/reduction system is *expansion terminating* if it is expansion terminating at each $n \in N$.

An $n \in N$ is in *expanded normal form* in case n is a reduced normal form and there are only reducible expansions leading from n. This equivalently means:

Lemma B.1. *$n \in N$ is an expanded normal form if and only if n is a reduced normal form with an expansion bound* $\mathrm{bnd}_{\mathscr{X}}(n) = 0$.

Proof. If the expansion bound exists and is zero, then certainly n is in expanded normal form. To show the converse consider a rewriting chain v_0 leading from an expanded normal form n which has a first irreducible expansion, $n_1 \xrightarrow{x} n_2$. The domain of this expansion step has reduced normal form n. This gives a divergence:

$$\begin{tikzcd} n_1 \arrow[r, "r"] \arrow[d, "x"'] & \mathrm{nf}_{\mathscr{R}}(n_1) \arrow[d, "v"] \\ n_2 \arrow[r, "r'"'] & n' \end{tikzcd}$$

Note that r' can contain no expansions and v can contain only one as the system is $\mathscr{X}$-reducing. Moreover, the expansion of v (if there is one) must occur immediately as there are no reductions of $\mathrm{nf}_{\mathscr{R}}(n_1)$. An expansion of an expanded normal form is always reducible, thus the first irreducible expansion of v_0 must be reducible! $\square$

If we wish to use an expansion/reduction system for deciding equality under the relation $(\mathscr{X} \cup \mathscr{R})^=$ then we should like to have a unique expanded normal form in each equivalence class. Such a system is called *expansion normalizing*. In general we have:

Proposition B.2. *An expansion/reduction system $\mathscr{X}/\mathscr{R}$ has at most one expanded normal form in each $(\mathscr{X} \cup \mathscr{R})^=$-equivalence class.*

Proof. Suppose that n_1 and n_2 are expanded normal forms in the same equivalence class, then, by confluence, there must be a convergence $n_1 \xrightarrow[*]{v_1} n' \xrightarrow[*]{v_2} n_2$. Further, we may assume that n' is a reduced normal form. Examining v_1, as the expansion bound of n_1 is zero, every expansion must be reducible. This means that n_1 and n' have the same reduced normal form. As they are both reduced normal forms they are equal. Similarly $n_2 = n'$. $\square$

We shall say that a rewrite chain v is *expansion normalized* if every expansion step of v is irreducible and has as its domain a reduced normal form.

Lemma B.3. *In an expansion/reduction system $\mathscr{X}/\mathscr{R}$, given any rewriting chain $n_1 \xrightarrow[*]{v} n_2$ there is an $\mathscr{X}$-normalized rewriting chain v' such that*

$$\begin{tikzcd} n_1 \arrow[r, "v"] \arrow[dr, "v'"'] & n_2 \arrow[d, "r_2"] \\ & \mathrm{nf}_{\mathscr{R}}(n_1) \end{tikzcd}$$

with $\mathrm{len}_{\mathscr{X}/\mathscr{R}}(v) = \mathrm{len}_{\mathscr{X}/\mathscr{R}}(v') = \mathrm{len}_{\mathscr{X}}(v')$.

Proof. We may argue by induction on the number of irreducible expansion steps in v. If there are none v' is just the reduction to normal form. If there is at least one let us examine the last,

$$n_1 \xrightarrow{v_1} m_1 \xrightarrow{x} m_2 \xrightarrow{v_2} n_2.$$

By considering the divergence of this expansion with the reduction of m_1 to normal form we obtain

$$\begin{tikzcd} m_1 \arrow[r, "x"] \arrow[d, "r_1"'] & m_2 \arrow[r, "v_2"] \arrow[d, "r'_2"'] \arrow[dr, "r'"] & n_2 \arrow[d, "r_2"] \\ \mathrm{nf}_{\mathscr{R}}(m_1) \arrow[r, "v'_1"'] & m'_2 \arrow[r, "r'_2"'] & \mathrm{nf}_{\mathscr{R}}(n_2) \end{tikzcd}$$

where as the square is $\mathscr{X}$-reducing v'_1 must have at most one expansion, which moreover must be the first reduction. In this manner we have normalized the tail and our induction hypothesis allows us to normalize the head as $v_1; r_1$ has one less irreducible expansion. $\square$

This means that the expansion bound, in an expansion-terminating system, is reached by the normalized rewritings. In addition, it means that following any expansion normalized rewriting must lead to an expanded normal form.

Proposition B.4. *In an expansion terminating expansion-reduction rewriting system $\mathscr{X}/\mathscr{R}$ each $(\mathscr{X} \cup \mathscr{R})^{=}$-equivalence class has a unique expanded normal form. Furthermore, each $n \in N$ has a rewriting chain $n \xrightarrow{v} \mathrm{nf}_{\mathscr{X}/\mathscr{R}}(n)$ which is expansion normalized.*

An expansion/reduction system is *expansion promoting* if whenever $n_1 \xrightarrow{r;x} n_2$, where $r \in \mathscr{R}^*$ and x is an expansion, then there is an $n_1 \xrightarrow{x';r'} n_2$, where $r' \in \mathscr{R}^*$ and x' is an expansion. In an $\mathscr{X}$-promoting system clearly there is another possible "normalized form" for the reductions: where all the (irreducible) expansions are done prior to reduction. Although a small observation this provides yet another approach to finding the expanded normal forms when this expansion promoting ability is present.

There remains one further, and quite crucial, observation about expansion/reduction systems:

Theorem B.5. *$\mathscr{X}/\mathscr{R}$ is an expansion/reduction system if and only if $\mathscr{R}$ is terminating and $\mathscr{X} \cup \mathscr{R}$ is locally confluent and $\mathscr{X}$-reducing.*

Proof. The only difficulty is establishing that local confluence and $\mathscr{X}$-reducing suffice. To prove this it is necessary to establish a series of facts:

1. By induction on $\mathrm{bnd}_{\mathscr{R}}(n)$ any divergence of the form $n_1 \xrightarrow{x} n \xrightarrow[*]{r} n_2$ (where $x \in \mathscr{X}$ and $r \in \mathscr{R}^*$) there is an $\mathscr{X}$-reducing convergence. When $\mathrm{bnd}_{\mathscr{R}}(n) = 0$, r must be empty and the result is immediately true. For $\mathrm{bnd}_{\mathscr{R}}(n) > 1$, if r has no steps there is nothing to prove, otherwise $\mathrm{len}(r) > 1$, that is $r = r_1; r'$, and the induction step is given (for the case when the expansion is not reduced: when it is reduced the result is immediate) by

$$\begin{tikzcd}
n \arrow[rrr, "x"] \arrow[d, "r_1"'] & & & n_1 \arrow[d] \\
n' \arrow[r, "r_{21}"] \arrow[d, "r'"'] \arrow[dr, phantom, "\text{confl}"] & m_1 \arrow[r, "x'"] \arrow[d] \arrow[dr, phantom, "\text{ind}"] & m_2 \arrow[r, "r_{12}"] \arrow[d] \arrow[dr, phantom, "\text{confl}"] & n'_1 \arrow[d] \\
n_2 \arrow[r, "r'_{21}"'] & m'_1 \arrow[r, "v"'] & m'_2 \arrow[r] & n'
\end{tikzcd}$$

where $\mathrm{bnd}_{\mathscr{R}}(m_1) < \mathrm{bnd}_{\mathscr{R}}(n)$.

2. Now consider a divergence of the form $n_1 \xrightarrow[*]{v} n \xrightarrow[*]{r} n_2$, where $v \in (\mathscr{X} \cup \mathscr{R})^*$ and $r \in \mathscr{R}^*$, we argue that there is an $\mathscr{X}$-reducing convergence for this by induction on $\mathrm{len}_{\mathscr{X}}(v)$.

When $\mathrm{len}_{\mathscr{X}}(v) = 0$ the convergence is given by the confluence of $\mathscr{R}$. Otherwise we tackle the first $\mathscr{X}$-step in v:

$$\begin{tikzcd}
n \arrow[r, "r_1"] \arrow[d, "r'"'] \arrow[dr, phantom, "\mathrm{confl}"] & n_{11} \arrow[r, "x"] \arrow[d, "r'"'] \arrow[dr, phantom, "(1)"] & n_{12} \arrow[r, "v_1"] \arrow[d, "r''"'] \arrow[dr, phantom, "\mathrm{ind}"] & n_1 \arrow[d, "r'''"] \\
n_2 \arrow[r, "r'_1"'] & n'_{11} \arrow[r, "v'"'] & n'_{12} \arrow[r, "v'_1"'] & n'
\end{tikzcd}$$

and note that v_1 has smaller $\mathscr{X}$-length.

3. Now consider a divergence of the form $n_1 \xrightarrow[*]{v_1} n \xrightarrow[*]{v_2} n_2$ where $v_1, v_2 \in (\mathscr{X} \cup \mathscr{R})^*$ with $\mathrm{len}_{\mathscr{X}}(v_i) \leq 1$ for $i = 1, 2$. We prove by induction on the reduction bound of n that this has a convergence which is $\mathscr{X}$-reducing.

The only case which presents any difficulty (using the confluence of $\mathscr{R}$ and (2) above) is when $\mathrm{len}_{\mathscr{X}}(v_1) = \mathrm{len}_{\mathscr{X}}(v_2) = 1$. Furthermore, when the $\mathscr{X}$-step is first in both v_1 and v_2 by using the local confluence we have

$$\begin{tikzcd}
n \arrow[r, "x_{11}"] \arrow[d, "x_{21}"'] \arrow[dr, phantom, "\mathrm{loc}"] & n_{11} \arrow[r, "r_{11}"] \arrow[d, "v_{21}"] \arrow[dr, phantom, "(2)"] & n_1 \arrow[d, "v'_{21}"] \\
n_{21} \arrow[r, "v_{11}"] \arrow[d, "r_{22}"'] \arrow[dr, phantom, "(2)"] & n'' \arrow[r, "r'_{11}"] \arrow[d, "r'_{22}"'] \arrow[dr, phantom, "\mathrm{confl}"] & n'_{21} \arrow[d, "r''_{22}"] \\
n_2 \arrow[r, "v'_{11}"'] & n_{12} \arrow[r, "r''_{11}"'] & n'
\end{tikzcd}$$

which easily provides a convergence of the desired form.

Now when the reduction bound of n is zero the above argument suffices to give a convergence of the desired form. Suppose now that $\mathrm{bnd}_{\mathscr{R}}(n) > 1$ and the $\mathscr{X}$-step does not occur first in both v_1 and v_2. Then we have

$$\begin{tikzcd}
n \arrow[r, "r_{11}"] \arrow[d, "r_{21}"'] \arrow[dr, phantom, "\mathrm{confl}"] & n_{11} \arrow[r, "v_{11}"] \arrow[d, "r'_{21}"] \arrow[dr, phantom, "(2)"] & n_1 \arrow[d, "r''_{21}"] \\
n_{21} \arrow[r, "r'_{11}"] \arrow[d, "v_{22}"'] \arrow[dr, phantom, "(2)"] & n'' \arrow[r, "v'_{11}"] \arrow[d, "v'_{22}"'] \arrow[dr, phantom, "\mathrm{ind}"] & n'_{21} \arrow[d, "v''_{22}"] \\
n_2 \arrow[r, "r''_{11}"'] & n_{12} \arrow[r, "v''_{11}"'] & n'
\end{tikzcd}$$

where n'' has a lower reduction bound.

The proof can now be completed for an arbitrary divergence $n_1 \xleftarrow[*]{v_1} n \xrightarrow[*]{v_2} n_2$ by an induction on $\mathrm{len}_{\mathscr{X}}(v_1) + \mathrm{len}_{\mathscr{X}}(v_2)$. If this is zero the confluence of $\mathscr{R}$ gives the result.

The only difficult case is when both these $\mathscr{X}$-lengths are greater than 1. In this case we may use (3) above to remove a single $\mathscr{X}$-step from each chain:

$$\begin{tikzcd}
n \arrow[r, "v_{11}"] \arrow[d, "v_{21}"'] \arrow[dr, phantom, "(3)"] & n_{11} \arrow[r, "v_{12}"] \arrow[d, "v'_{21}"] \arrow[dr, phantom, "\text{ind}"] & n_1 \arrow[d, "v''_{21}"] \\
n_{21} \arrow[r, "v'_{11}"] \arrow[d, "v_{22}"'] \arrow[dr, phantom, "\text{ind}"] & n'' \arrow[r, "v'_{11}"] \arrow[d, "v'_{22}"'] \arrow[dr, phantom, "\text{ind}"] & n'_{21} \arrow[d, "v''_{22}"] \\
n_2 \arrow[r, "r''_{11}"'] & n_{12} \arrow[r, "v''_{11}"'] & n'
\end{tikzcd}$$

Notice now that the problems based at n_{11}, n_{22}, and n'' are now smaller. $\square$

If the system is given by a set of redex–reduct pairs, and the convergences for non-overlapping divergences are $\mathscr{X}$-reducing, then showing that one has an expansion/reduction system amounts to checking for the termination of the reductions and that critical divergences have reducing convergences. It is worth noting that, while the condition on non-overlapping divergences is certainly valid for linear systems (and, in particular, for the nets we consider), it is by no means automatic. In particular the η/β expansion/reduction system for the λ-calculus violates the expansion reducing condition for non-overlapping redexes because the β rule can duplicate η redexes. As expansion/reduction systems provide a desirable framework, it is often worth recasting examples, if possible, so that they fit. A way to bring the typed λ-calculus to heel is described in [10] where they replaced the η by a "parallel η" rule.

To be assured of expanded normal forms it suffices to, in addition, show expansion-termination. When the system is expansion promoting it suffices to show that the number of irreducible expansions which can be applied directly to a term is bounded.

B.3. Expansion/reduction modulo equations

We shall be mainly concerned with a slight generalization of expansion/reduction systems in which there is a third system of rewrites involved, denoted $\mathscr{E}$. These we shall regard as equations which can be used in either orientation.

We start by generalizing all our previous definitions so that they can be read as modulo equations. Thus $\mathscr{R}$ is *$\mathscr{E}$-terminating at $n \in N$* in case there is a (least) $\mathrm{bnd}_{\mathscr{R}}(n) \in \mathbb{N}$ which bounds the number of reductions in every $\mathscr{R} \cup \mathscr{E} \cup \mathscr{E}^{\mathrm{op}}$-rewrite chain $n \xrightarrow[*]{v} n'$, that is $\mathrm{bnd}_{\mathscr{R}}(n) \geq \mathrm{len}_{\mathscr{R}}(v)$. As before, $n \in N$ is an $\mathscr{E}$-normal form if and only if $\mathrm{bnd}_{\mathscr{R}}(n) = 0$. $\mathscr{R}$ is *$\mathscr{E}$-confluent* in case for every $\mathscr{R} \cup \mathscr{E} \cup \mathscr{E}^{\mathrm{op}}$-divergence there is a corresponding convergence. We shall say that $\mathscr{R}$ is an *$\mathscr{E}$-reduction system* when $\mathscr{R}$ is $\mathscr{E}$-terminating and $\mathscr{E}$-confluent.

Notice the pattern of these definitions: if we form the equivalence relation $\mathscr{E}^=$ and quotient N, then the induced rewriting $\mathscr{R}$ on $N/\mathscr{E}^=$ (note that this may not be a relation even if $\mathscr{R}$ had been but will be a directed graph) has the given property if and only if $\mathscr{R}$ has it modulo $\mathscr{E}$. This is a convention which seems sensible to follow.

An *expansion/reduction system modulo equations*, $\mathscr{X}/\mathscr{R}$ modulo $\mathscr{E}$ on N consists of an $\mathscr{E}$-reduction system $\mathscr{R}$ together with a system of expansions $\mathscr{X}$ such that $\mathscr{X} \cup \mathscr{R}$ is $\mathscr{E}$-confluent and $\mathscr{X}$-reducing.

By moving to $N/\mathscr{E}$ and back it is easy to see that we have the analogue of the last sections' results. In particular, B.5 becomes:

Theorem B.6. *$\mathscr{X}/\mathscr{R}$ modulo $\mathscr{E}$ is an expansion/reduction system modulo equations if and only if $\mathscr{R}$ is $\mathscr{E}$-terminating and $\mathscr{X} \cup \mathscr{R}$ is locally $\mathscr{E}$-confluent and $\mathscr{X}$-reducing.*

For an $\mathscr{E}$-reduction system the normal form of $n \in N$ is an $\mathscr{E}^=$-equivalence class $\mathrm{nf}_{\mathscr{R}}(n)$. Similarly, an $\mathscr{E}$-expanded form is an $\mathscr{E}$-reduced form, all of whose $\mathscr{E}^=$ equivalent forms have only $\mathscr{E}$-reducible expansions. The uniqueness of these forms will follow the same pattern as before.

Notice that the above proposition only partially localizes the problem of determining whether one is holding such a system: it does not remove the necessity of having to handle arbitrary sequences of equalities. In general this is neither desirable nor, indeed, even feasible. Furthermore, the process of normalizing is much more complex: we may have to continually shift between $\mathscr{E}^=$-equivalent forms in the process of doing an expansion reduction.

We say that a reduction/expansion system modulo equations, $\mathscr{X}/\mathscr{R}$ modulo $\mathscr{E}$, is *equation reducing* in case it is $\mathscr{X} \cup \mathscr{R} \cup \mathscr{E} \cup \mathscr{E}^{\mathrm{op}}$-confluent and $\mathscr{E}$-reducing (as well as $\mathscr{X}$-reducing).

Theorem B.7. *$\mathscr{X}/\mathscr{R}$ modulo $\mathscr{E}$ is an expansion/reduction system modulo equations which is $\mathscr{E}$-reducing if and only if $\mathscr{R}$ is terminating and $\mathscr{X} \cup \mathscr{R} \cup \mathscr{E} \cup \mathscr{E}^{\mathrm{op}}$ is locally confluent and $\mathscr{E}$-and $\mathscr{X}$-reducing.*

Again the problem is to show that the more local condition suffices. Fortunately, this is now easy as the system is $\mathscr{X} \cup \mathscr{E} \cup \mathscr{E}^{\mathrm{op}}$-reducing over $\mathscr{R}$.

This means that the problem of showing that we have an expansion/reduction system modulo a set of equations can be reduced to checking that $\mathscr{R}$ is terminating (no mention of $\mathscr{E}$ here) and the local critical divergences have convergences satisfying the required reducing properties. If one wants to secure expanded normal forms (modulo the equations) it suffices, in general, to provide an $\mathscr{E}$-expansion bound. When the system is $\mathscr{E}$-reducing this is simply an expansion bound for $\mathscr{X}/\mathscr{R}$ and the notion of a normalized rewrite chain is exactly as before – no equations are allowed.

If the system is expansion $\mathscr{E}$-terminating then each $n \in N$ will have an expanded $\mathscr{E}$-normal form. This means that the resulting form is one of a number of possible $\mathscr{E}$-equivalent forms. To complete a decision procedure one needs, as a last step, a procedure for determining $\mathscr{E}$-equivalence.

References

[1] V.M. Abrusci, Phase semantics and sequent calculus for pure noncommutative classical linear propositional logic, J. Symbolic Logic 56 (1991) 1403–1451.

[2] G. Bellin and J. van de Wiele, Empires and kingdoms in MLL^-, in: J.-Y. Girard, Y. Lafont and L. Regnier, Eds., Advances in Linear Logic, London Math. Soc. Lecture Notes Series, Vol. 222 (1995) 249–270.

[3] R.F. Blute, Proof nets and coherence theorems, in: Categories and Computer Science, Paris, 1991, Lecture Notes in Computer Science, Vol. 530 (Springer,Berlin, 1991).

[4] R.F. Blute, Linear logic, coherence and dinaturality, Theoret. Comput. Sci. 115 (1992) 3–42.

[5] R.F. Blute, J.R.B. Cockett and R.A.G. Seely, ! and ? for weakly distributive categories: storage as tensorial strength, preprint, McGill University, 1992; to appear in Math. Struct. Comput. Sci.

[6] J.R.B. Cockett and R.A.G. Seely, Weakly distributive categories, in M.P. Fourman, P.T. Johnstone, and A.M. Pitts, Eds., Applications of Categories to Computer Science, London Math. Soc. Lecture Note Series Vol. 177 (1992) 45–65; expanded version to appear in J. Pure Appl. Algebra.

[7] V. Danos, La logique linéaire appliquée à l'étude de divers processus de normalisation et principalement du λ-calcul, Doctoral Dissertation, Université de Paris, 1990.

[8] V. Danos and L. Regnier, The structure of multiplicatives, Arch. Math. Logic 28 (1989) 181–203.

[9] A. Fleury and C. Retoré, The mix rule, Math. Struct. Comput. Sci. 4 (1994) 273–285.

[10] N. Ghani and C.B. Jay, The virtues of η-expansion, to appear in J. Funct. Programming.

[11] J.-Y. Girard, Linear logic, Theoret. Comput. Sci. 50 (1987) 1–102.

[12] J.-Y. Girard, Linear logic: a survey, preprint, 1992.

[13] J.-Y. Girard, A. Scedrov, and P.J. Scott, Bounded linear logic, Theoret. Comput. Sci. 97 (1992) 1–66.

[14] C.B. Jay, The structure of free closed categories, J. Pure Appl. Algebra 66 (1990) 271–285.

[15] A. Joyal and R. Street, The geometry of tensor calculus I, Adv. in Math. 88 (1991) 55–112.

[16] G.M. Kelly, On MacLane's conditions for coherence of natural associativities, commutativities, etc., J. Algebra 1 (1964) 397–402.

[17] G.M. Kelly and S. MacLane, Coherence in closed categories, J. Pure and Appl. Algebra 1 (1972) 97–140.

[18] Y. Lafont, From proof nets to interaction nets, in: J.-Y. Girard, Y. Lafont and L. Regnier, Eds., Advances in Linear Logic, London Math. Soc. Lecture Notes Series, Vol. 222 (1995) 225–248.

[19] J. Lambek, Deductive Systems and Categories II, Lecture Notes in Mathematics Vol. 87 (Springer, Berlin, 1969).

[20] J. Lambek, From categorial grammar to bilinear logic, in: K. Došen and P. Schroeder-Heister, Eds., Substructural Logics, Proc. Tübingen Conf., 1990.

[21] J. Lambek, and P.J. Scott, Introduction to Higher-Order Categorical Logic, Cambridge Studies in Advanced Mathematics, Vol. 7 (Cambridge Univ. Press, Cambrdige, 1986).

[22] M. Laplaza, Coherence for distributivity, Lecture Notes in Mathematics, Vol. 281 (Springer, Berlin, 1972) 29–66.

[23] G. Mints, Closed categories and the theory of proofs, J. Sov. Math. 15 (1981) 45–62.

[24] D. Prawitz, Natural Deduction, (Almqvist and Wiksell, Uppsala, 1965).

[25] R.A.G Seely, Linear logic, $*$-autonomous categories and cofree coalgebras, in: J. Gray and A. Scedrov, Eds., Categories in Computer Science and Logic, Contemporary Mathematics, Vol. 92 (Amer. Mathematical Soc., Providence, RI, 1989).

[26] S. Soloviev, On natural transformations of distinguished functors and their superpositions in certain closed categories, J. Pure Appl. Algebra 47 (1987) 181–204.

[27] M.E. Szabo, Polycategories, Comm. Algebra 3 (1975) 663–689.

[28] T.H. Trimble, Linear logic, bimodules, and full coherence for autonomous categories, Doctoral Dissertation, Rutgers University, 1994.

[29] A. Troelstra, Lectures on Linear Logic, CSLI Lecture Notes, Vol. 29 (CSLI, 1992).