Reduction as Deduction

John Darlington  Yi-ke Guo  Martin Kohler
Department of Computing
Imperial College
180 Queen’s Gate, London SW7 2BZ, U.K.
E-mail: jd, yg, mk@doc.ic.ac.uk

Functional Languages Implementation Workshop. Paper 10

1 Introduction

In this paper, we present a logical encoding of reduction strategies of lambda calculus in terms of linear logic. First of all, we present a fragment of intuitionistic linear logic to formalise the behaviours of concurrent computation. The logic, which can be regarded as a process calculus, is then used as a meta-language to specify the behaviors of various reduction strategies of the calculus. The encoding maps a lambda expressions into linear logic formulae by taking into account its computational behaviours. With such encoding, reductions are systematically mapped into a uniform framework of linear deduction.

2 A Linear Logic Model of Computation

Linear logic was proposed by Girard as a logic for modelling computational systems [Gir87]. In linear logic, the structural rules for weakening and contraction are dropped. Therefore, it is not possible to copy or discard arbitrary formulae during a deduction. Logical formulae become resources which may be consumed by deduction. This property makes the logic ideal for specifying computational systems, and particularly concurrent computation. In the appendix, we present linear logic by its sequent system where \( \rightarrow \) is adopted as a basic logical operation rather than a derived operation. A comprehensive description of linear logic can be found in [Gir87].

Consider the sequent \( \Gamma \vdash \Delta \) as consuming the resource of \( \Gamma \) to meet the requirement of \( \Delta \). A linear proof can be regarded as a resource construction/consumption procedure where the root of the proof tree describes the initial state of a computation which proceeds by expanding the frontier of the partially constructed tree. The collection of leaves presents the current state of the computation. Following this view, the multiset of formulae \( \Gamma \) in a sequent \( \Gamma \vdash \Delta \) is viewed as a multiset of resources. These resources may be transformed into new resources by applying the inference rules (left-rules). Since there has no sequentiality imposed on transforming these resources, the formulae in \( \Gamma \) can be regarded as concurrent processes. Logic operators (connectives and quantifiers) become combinators of the processes. Regarding each sequent as a state of the computation, the operational semantics of these combinators are provided by identifying state transitions of the computation with the inferences of the logic. The proof tree as a whole records the history of computation.

Thus, we can select an appropriate fragment of the logic as a process calculus. From a logical point of view, any statement of the calculus is a logical formula, whereas, from a computational point of view, a formula is a process. The concurrent computation of the language is modelled by deduction in the logic. The following analogies arise:

<table>
<thead>
<tr>
<th>Logic</th>
<th>Computation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formula</td>
<td>Process</td>
</tr>
<tr>
<td>Sequent</td>
<td>Configuration</td>
</tr>
<tr>
<td>Deduction</td>
<td>Concurrent Computation</td>
</tr>
</tbody>
</table>
Following this principle of “formulae as processes, proofs as computations”, basic constructions for concurrent computation can be straightforwardly related to an operational interpretation of linear deduction. A detailed description can be found in [Guo03].

We will now present a logical process calculus as a meta language for the operational behaviour of the λ-calculus. It is based on the intuitionistic linear logic [Abr90]. We first present the syntax of the language.

Let $A, x, t, c$ be the syntactical variables ranging over agents, channel symbols, terms and conditional expressions respectively. The language has the following syntax:

$$
\begin{align*}
A & \triangleq 1 \quad \text{unit} \\
& \mid x(t) \quad \text{atomic communication} \\
\text{mid } c & \rightarrow A \quad \text{condition} \\
& \mid \forall \bar{x} . m \rightarrow A \quad \text{communicating method} \\
& \mid A \otimes A \quad \text{parallel composition} \\
& \mid A \& A \quad \text{choice} \\
& \mid \exists x . A \quad \text{hiding} \\
& \mid !A \quad \text{storing}
\end{align*}
$$

where the syntactical variable $m$ ranges over a set of templates defined as

$$
\begin{align*}
m & \triangleq x(t) \quad \text{receiver} \\
& \mid m \otimes m \quad \text{multiple receiver}
\end{align*}
$$

A condition $c$ in an implication $!c \rightarrow A$ is generally a preinterpreted formula. We assume that a built-in theory to define the meaning of the symbols of $c$.

Communication methods are generally of the form

$$
(\forall \bar{x} . p_1(\bar{x}_1) \otimes \ldots \otimes p_n(\bar{x}_n) \rightarrow A).
$$

They are also called multiheaded clauses and the atomic propositions $p_i(\bar{x}_i)$ are called the heads of the clause. Computation is to take these methods as rewriting rules to reduce a multiset of formulae into a store of consistent bindings. Each application of a method forms a multiset rewriting step.

The concurrency of the computation is reflected by using the multiplicative conjunction $\otimes$ as an operator for parallel composition. Its deduction rule

$$
\frac{\Gamma, A, B \vdash \Delta}{\Gamma, A \otimes B \vdash \Delta}
$$

can be understood as decomposing the resource $A \otimes B$ into the resources $A$ and $B$ in the context $\Gamma$.

Consider the deduction rule $(\& \vdash)$:

$$
(\& \vdash 1) \quad \frac{\Gamma, A \vdash \Delta}{\Gamma, A \& B \vdash \Delta} \quad (\& \vdash 2) \quad \frac{\Gamma, B \vdash \Delta}{\Gamma, A \& B \vdash \Delta}
$$

The deduction can be regarded as choosing one of the component of $A \& B$ to verify the sequent within the context $\Gamma$. Since computation is to search for a successful proof, the choice may be influenced by other formulae in $\Gamma$ (i.e. the environment). This corresponds to the “external choice” in concurrent programming [BG90]. Thus, don’t care non-determinism (committed-choice), for which the arrival nondeterminism is a special case, can be uniformly modelled in terms of additive conjunction.

Communication between processes is realized by passing messages, as outlined in the following deduction:

$$
\frac{\Gamma, B, A[t_1/y] \vdash \Delta}{\Gamma, B, x(t_1), A[t_1/y] \vdash \Delta} \quad (\rightarrow \vdash) \\
\frac{\Gamma, B, x(t_1), A[t_1/y] \vdash \Delta}{\Gamma, B, x(t_1), \forall y . x(y) \rightarrow A \vdash \Delta} \quad (\forall \vdash) \\
\frac{\Gamma, B \otimes x(t_1), \forall y . x(y) \rightarrow A \vdash \Delta}{\Gamma, B \otimes x(t_1) \vdash \Delta} \quad (\otimes \vdash)
$$

The formula $x(t_1)$ can be understood as the sending of a message $(t_1)$ via a channel $x$. Thus, the formula $B \otimes x(t_1)$ sends the message $x(t_1)$ and then behaves like $B$. The formula $\forall y . x(y) \rightarrow A$ receives the message
$x(t_1)$ and behaves like $A[t_1/x]$. In general, synchronization between processes in the calculus is modelled by linear modus ponens. Consider the following deduction rule for linear implication:

$\Gamma_1 \vdash A, \Delta_1 \quad \Gamma_2, B \vdash \Delta_2$

$\Gamma_1, \Gamma_2, A \rightarrow B \vdash \Delta_1, \Delta_2$

A special case of the inference rule is that the succedent of the lower sequent (the global requirement of computation) isn’t split during deduction. That is:

$\Gamma_1 \vdash A \quad \Gamma_2, B \vdash \Delta$

$\Gamma_1, \Gamma_2, A \rightarrow B \vdash \Delta$

The deduction step can be interpreted as transforming the current state, represented as $\Gamma_1, \Gamma_2, A \rightarrow B \vdash \Delta$, to the state $\Gamma_2, B \vdash \Delta$ whenever the local requirement condition $A$ is met by $\Gamma_1$ (i.e. $\Gamma_1 \vdash A$). Then the deduction, called linear modus ponens, can be presented as:

$\Gamma_2, B \vdash \Delta$

$\Gamma_1, \Gamma_2, A \rightarrow B \vdash \Delta$

if $\Gamma_1 \vdash A$

With this interpretation, linear modus ponens forms a uniform synchronisation mechanism where the implication $A \rightarrow B$ can be regarded as a process which will become $B$ whenever the environment contains enough resources to meet the synchronisation condition (local requirement) represented by the premise $A$. Synchronisation is logically modelled as verifying the entailment between the resource $\Gamma_1$ and the local requirement $A$. In the case of communication, the local requirement $A$ can be regarded as a message pattern waiting for matching the corresponding messages. Thus, linear modus ponens provides a concise means to model asynchronous communication.

The condition $c$ in a formula $!c \rightarrow A$ states the synchronisation condition between the environment and $A$. $A$ can proceed only when $c$ is justified with respect to recently computed variable bindings. The effect can be written as $\vdash_S c$ where $S$ denotes the built-in theory asserting the truth of the formula $c$ (Thus, the theory should be represented in classical logic). The modality $!$ allows the formula $c$ to be accepted by the built-in theory in classical logic.

Variables local to a process can be introduced via existential quantification, because an existentially quantified variable can not be free outside the scope of $\exists$:

$\Gamma, A \vdash x$ is not free in $\Gamma$

The following derived rule, called $CD$ because it combines contraction and dereliction, indicates that $!A$ can be understood as storing the process $A$:

$!A, A, \Delta \vdash \Gamma$

$!A, \Delta \vdash \Gamma$

$CD$ – Rule

The inference is read as “making a copy of $A$ and putting it into the environment”. In the calculus, we use $!$ to store resources.

The operational model of the process calculus is given as a special form of deduction in linear logic. It can be presented as a transition system where a configuration $\Gamma$ is a multiset of agents. We use $Con$ to denote the set of all configurations. A derivation relation $\vdash \subseteq Con \times Con$ represents the transition between configurations, where $\vdash S c$ denotes the free variables in the agents $\Gamma$.

The rules of the transition system are defined as follows:

$\Gamma, 1 \vdash \Gamma$

$\Gamma, A \otimes B \vdash \Gamma, A, B$

$\Gamma, A \& B \vdash \Gamma, A$

$\Gamma, A \& B \vdash \Gamma, B$

$\Gamma, !c \rightarrow A \vdash \Gamma, A$ if $\vdash_S c$

$\Gamma, x_1(t_1), x_2(t_2), \ldots, x_n(t_n), (\forall y. \otimes_{i=1}^{n} x_i(y_i) \rightarrow A) \vdash \Gamma.A[t_i/y_i]$

---

$^1$The interpretation imposes a sequentiosity of verifying the left branch first.
\[ \Gamma, \exists x.A \Rightarrow \Gamma.A \quad x \not\in \varnothing(\Gamma) \]

\[ \Gamma, \top \Rightarrow \Gamma, \top, A \]

For any initial formula \( A \), the initial configuration is \( \Gamma_0 : \{ A \} \). The computation transforms the initial configuration into a simpler system \( \Gamma_n \) such that no more communication is possible. The transition rules are deduction steps in linear logic. As proved in [DG92], the transition system realizes a special form of proofs in linear logic. The computation is complete in the sense that it constructs a proof that is canonical for all possible evaluations of \( A \). The soundness and completeness results can be summarized in the following theorem where we use \( \Gamma.C \) to denote the provability in linear logic\(^2\).

**Theorem 2.1** For any agent \( A \), a term \( t \) and \( x \in \varnothing(A) \), \( A \vdash_C x(t) \) if and only if there exists a derivation \( \Gamma_0 : \{ A \} \Rightarrow^* \Gamma_n \) such that \( \Gamma_n \vdash_C x(t) \)

### 3 Encoding Call by Name Reduction in Linear Logic

In this section, we present a simple linear logic encoding of the call by name reduction model of \( \lambda \)-calculus by encoding each expression into a linear logic formula. Reduction in \( \lambda \)-expressions thus corresponds to deduction of the formula. With this encoding, a call by name evaluation model is presented within the linear deduction framework. The encoding of call by name lambda reduction is:

\[
\begin{align*}
[x]_y^o &= x(y) \\
[\lambda x.M]_y^o &= \forall x, y. (z(x, y) \rightarrow [M]_y^o) \\
[(M \cdot N)]_y^o &= \exists x, y. ([M]_y^o \otimes x(y, z) \otimes [N]_y^o) \\
& \quad \text{if } x, y \text{ is not free in } M, N \\
[x := N]_y^o &= \forall \forall m. x(m) \rightarrow [N]_m^o
\end{align*}
\]

According to the concurrent semantics of the calculus, the atomic formula \( x(y) \) is read as "sending the channel name \( y \) on the channel \( x \)". The atomic formula \( x(y, z) \) is understood as the transmission of a tuple \((y, z)\) along the channel \( x \). The binding \( y := \overline{N} \) can be interpreted as adding an entry into an environment, which is built up during the computation and contains a set of communication methods.

The intuition underlying our transformation is to model \( \lambda \)-expressions directly as communicating processes, i.e. the formula \( [e]_y^o \) is interpreted as a process which computes \( e \) and associates the channel \( x \) with the result of the computation.

**Function as Servers**: The translation of functions (\( \lambda \)-abstractions)

\[ [\lambda x.M]_y^o = \forall x, y. z(x, y) \rightarrow [M]_y^o \]

 associates the abstraction with the channel \( z \) and translates it into a communication method \( \forall x, y. z(x, y) \rightarrow [M]_y^o \). This formula behaves like a server waiting for a message to arrive along the channel \( z \) indicating the name of the channel along which the argument of the function will be passed when the function is applied and the name of the channel to which the result will be sent. Thus, in this translation, function names, variables and communication channels are identified. This approach directly follows Milner's translation of the \( \lambda \)-calculus into the \( \pi \)-calculus [MPD89]. \( \beta \)-reduction corresponds to a communication which is realised by a combination of universal instantiation and linear modus ponens as indicated by the following deduction:

\[
\begin{align*}
\Gamma, [M[u/x]]_y^o \vdash \Delta \\
\Gamma, z(u, v), z(u, v) \rightarrow [M[u/x]]_y^o \vdash \Delta \\
\Gamma, z(u, v), (\forall z, y. z(x, y) \rightarrow [M]_y^o) \vdash \Delta
\end{align*}
\]

Following the identity of function names and communication channels, the formula \( z(u, v) \) can be understood as either sending the tupled channel names \((u, v)\) via the channel \( z \), or a function call formed by applying function \( z \) to \( u \) with channel \( v \) transmitting the result. Combined, they allow the deduction of a formula that represents the body of the original \( \lambda \)-abstraction, with the argument identifier \( x \) substituted by the argument name \( u \). This substitution creates a specific instance of the function body \( M \) which is equal in meaning to \([M[u/x]]_v^o\). The result of this specialised function body will therefore be computed using \( u \) and the result will be transmitted along the channel \( v \), as desired. The deduction shows that linear modus ponens provides a logical interpretation of \( \lambda \)-rewriting.

\(^2\)More precisely, we assume linear logic integrated with a built-in theory for justifying conditions.
Application as Parallel Composition: The translation of function application:

\[
[(MN)]_2^o = \exists x, y. ([M]_2^o \otimes [y := N]^o \otimes x(y, z))
\]

takes the behaviour of call by name reduction into account. Two intermediate names \(x, y\) are created for private channels that will be used as a function name and for the transmission of the argument, respectively. The formulae for the function and the argument expression are linked via parallel composition but the evaluation of the argument expression is suspended and stored in a "logical environment" until a demand is received. The third formula: \(x(y, z)\) "feeds" the name \(y\) of the argument into the function \(x\) and ensures that the result of the function application is sent onto the channel \(z\). Moreover, the argument \(N\) of the application is associated with the name \(y\) as its environment entry. It will only be evaluated if and when its value is required for the computation of the result of the application. The demand for evaluation is propagated by sending to the environment the channel on which the value of \(N\) is expected. Thus, an environment entry can be logically formalised as:

\[
[y := N]^o := ! (\forall m. y(m) \rightarrow [N]^o_m)
\]

which can be understood as a suspension for the expression \(N\), waiting for a demand transmitted on channel \(y\). The demand has the form of a message \(y(m)\) where \(m\) is the channel over which the result of \(N\) should be transmitted. The view of an environment as a set of suspensions is consistent with the notion of environments in the lazy functional programming context where a suspended computation is encoded as a stored closure. Note that whenever the argument is needed, it will be copied and be reevaluated. Thus, the encoding reflects the "call-by-name" mechanism for parameter passing.

The following translation theorem (theorem 3.0.1), which is proved in [Guo93], states the soundness of the encoding.

**Theorem 3.0.1 (Translation Theorem)** Let \(\epsilon_1, \epsilon_2\) be two lambda-expressions and the relation \(\rightarrow^\beta\) denote a call by name \(\beta\)-conversion step. Then:

\[
\epsilon_1 \rightarrow^\beta \epsilon_2 \Rightarrow [\epsilon_1]^o_2 \vdash_LL [\epsilon_2]^o_2
\]

4 A General Encoding of Lambda Reduction Models

A general encoding of lambda reduction models is presented as follows:

\[
\begin{align*}
[x]_2^o &= x(z) \quad \text{Name passing} \\
[(\lambda x.M)]_2^o &= \exists f, z(f) \otimes (\forall x, y. (f(x, y) \rightarrow [M]_2^o)) \quad \text{Abstraction as server} \\
[x := N]^o &= !(\forall m. x(m) \rightarrow [N]^o_m) \quad \text{Stored environment entry} \\
[(MN)]_2^o &= \exists x, y, u. ([N]^o_y \otimes (\forall v. u(v) \rightarrow [M]_2^o \otimes (\forall f. x(f) \rightarrow f(y, z) \otimes [y := v]^o))) \quad \text{Call by value reduction} \\
[(MN)]_2^o &= \exists x, y, u. ([M]_2^o \otimes [y := N]^o \otimes (\forall f. x(f) \rightarrow f(y, z))) \quad \text{Call by name continuation} \\
[(MN)]_2^o &= \exists x, y, u. ([M]_2^o \otimes [y := N]^o \otimes (\forall f, v. x(f) \otimes u(v) \rightarrow f(y, z) \otimes [y := v]^o))) \quad \text{Call by need reduction} \\
[(MN)]_2^o &= \exists x, y, u. ([M]_2^o \otimes [N]^o_u \otimes (\forall f, v. x(f) \otimes u(v) \rightarrow f(y, z) \otimes [y := v]^o)) \quad \text{Data flow Model}
\end{align*}
\]

References


A Linear Logic

Linear logic is presented in Figure 1 in the form of a Gentzen-style sequent calculus. Since the structural rules for weakening and contraction are abolished in linear logic, a collection of formulae in a deduction should be regarded as a multiset rather than a set. The feature of resource-consciousness results from the linearity of the proofs which requires that each assumption must be used exactly once. For a sequent $\Gamma_1 \ldots \Gamma_n \vdash \Delta_1 \ldots \Delta_m$ the assumptions $\Gamma_i$ can be viewed as resources and the conclusions $\Delta_j$ as requirements that have to be met by spending the given resources.

The linearity necessitates two different versions of conjunction: the multiplicative conjunction ($\text{tensor}$, $\otimes$) and the additive conjunction ($\text{with}$, $\&$). If the multiplicative conjunction $\otimes$ is used in an assumption $\Gamma_i$ of a sequent, then this reflects the fact that both components of the conjunction must make contributions to the proof derivation within the same environment. On the other hand, the use of the additive conjunction $\&$ reflects the fact that once the formula is used in the proof, either the first or the second component must be chosen for the further derivation. Dually, two versions of disjunction are required. The multiplicative disjunction ($\text{par}$, $\otimes$) causes the splitting of its environment, whereas additive disjunction ($\text{plus}$, $\oplus$) causes the environment to be copied for the derivation of both components.

Another important connective is linear implication ($\rightarrow$), which can be used to localize inference. It can be derived by the logical equivalence $A \rightarrow B = A^\perp \otimes B$ where $A^\perp$ is the linear negation of $A$. Each sequent $\Gamma_1 \ldots \Gamma_n \vdash \Delta_1 \ldots \Delta_m$ has the intended meaning that $\otimes_{i=1}^n \Gamma_i \rightarrow \otimes_{j=1}^m \Delta_j$ is valid.

The “of course” modality (!) is introduced to allow a controlled form of weakening and contraction in the language. Its dual (?) is called “why not”.

Due to the symmetry of linear logic, the same multiset of formulae can be treated either as assumptions, using conjunction, or as consequences, using disjunction. Here, we concentrate on the interpretation of formulae as resources and there only deal with the rules for the left-hand side of sequents.

For a comprehensive description of linear logic see [Gir87].
Axiom:

Cut:

\[
\frac{\Gamma_1 \vdash G \Delta_1 \quad \Gamma_2 \vdash G \Delta_2}{\Gamma_1 \cup \Gamma_2 \vdash G \Delta_1 \Delta_2}
\]

Structure Rule:

<table>
<thead>
<tr>
<th>Logic Rules</th>
<th>(Exchange Left)</th>
<th>(Exchange Right)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1 ⊢)</td>
<td>[\Gamma, A \Delta \vdash \Gamma, \Gamma, A \Delta ]</td>
<td>[\Gamma, \Gamma, A \Delta \vdash \Gamma, \Gamma, \Gamma, A \Delta ]</td>
</tr>
<tr>
<td>(0 ⊢)</td>
<td>[\Gamma, 0 \Delta \vdash \Gamma, \Gamma, 0 \Delta ]</td>
<td>[\Gamma, \Gamma, 0 \Delta \vdash \Gamma, \Gamma, \Gamma, 0 \Delta ]</td>
</tr>
<tr>
<td>(∥ ⊢)</td>
<td>[\Gamma, A \Delta \vdash \Gamma, A \Delta ]</td>
<td>[\Gamma, \Gamma, A \Delta \vdash \Gamma, \Gamma, A \Delta ]</td>
</tr>
<tr>
<td>(⊗ ⊢)</td>
<td>[\Gamma, A, B \Delta \vdash \Gamma, A \Delta ]</td>
<td>[\Gamma_1 \vdash A, A \Delta_1 \Delta_2 ]</td>
</tr>
<tr>
<td>(⊔ ⊢)</td>
<td>[\Gamma, A \Delta \vdash \Gamma, A \Delta ]</td>
<td>[\Gamma_1 \vdash A, A \Delta_1 \Delta_2 ]</td>
</tr>
<tr>
<td>(∨ ⊢)</td>
<td>[\Gamma, A \Delta \vdash \Gamma, A \Delta ]</td>
<td>[\Gamma_1 \vdash A, A \Delta_1 \Delta_2 ]</td>
</tr>
<tr>
<td>(∼ ⊢)</td>
<td>[\Gamma, A, \Delta \vdash \Gamma, A \Delta ]</td>
<td>[\Gamma_1 \vdash A, A \Delta_1 \Delta_2 ]</td>
</tr>
<tr>
<td>(Contraction !)</td>
<td>[\Gamma, A, \Delta \vdash \Gamma, A \Delta ]</td>
<td>[\Gamma_1 \vdash A, A \Delta_1 \Delta_2 ]</td>
</tr>
<tr>
<td>(Weakening !)</td>
<td>[\Gamma, A \Delta \vdash \Gamma, A \Delta ]</td>
<td>[\Gamma, A \Delta \vdash \Gamma, A \Delta ]</td>
</tr>
<tr>
<td>(Promotion ?)</td>
<td>[\Gamma, A \Delta \vdash \Gamma, A \Delta ]</td>
<td>[\Gamma, A \Delta \vdash \Gamma, A \Delta ]</td>
</tr>
<tr>
<td>(Dereliction !)</td>
<td>[\Gamma, A \Delta \vdash \Gamma, A \Delta ]</td>
<td>[\Gamma, A \Delta \vdash \Gamma, A \Delta ]</td>
</tr>
<tr>
<td>(∃ ⊢ *)</td>
<td>[\Gamma, A \Delta \vdash \Gamma, A \Delta ]</td>
<td>[\Gamma, A \Delta \vdash \Gamma, A \Delta ]</td>
</tr>
<tr>
<td>(∀ ⊢)</td>
<td>[\Gamma, A \Delta \vdash \Gamma, A \Delta ]</td>
<td>[\Gamma, A \Delta \vdash \Gamma, A \Delta ]</td>
</tr>
</tbody>
</table>

* x is not free in lower sequent.

Figure 1: Linear Logic