Elaborating Intersection and Union Types

Jana Dunfield
Max Planck Institute for Software Systems
Kaiserslautern and Saarbrücken, Germany
jd169@queensu.ca

Abstract
Designing and implementing typed programming languages is hard. Every new type system feature requires extending the metatheory and implementation, which are often complicated and fragile. To ease this process, we would like to provide general mechanisms that subsume many different features.

In modern type systems, parametric polymorphism is fundamental, but intersection polymorphism has gained little traction in programming languages. Most practical intersection type systems have supported only refinement intersections, which increase the expressiveness of types (more precise properties can be checked) without altering the expressiveness of terms; refinement intersections can simply be erased during compilation. In contrast, unrestricted intersections increase the expressiveness of terms, and can be used to encode diverse language features, promising an economy of both theory and implementation.

We describe a foundation for compiling unrestricted intersection and union types: an elaboration type system that generates ordinary $\lambda$-calculus terms. The key feature is a Forsythe-like merge construct. With this construct, not all reductions of the source program preserve types; however, we prove that ordinary call-by-value evaluation of the elaborated program corresponds to a type-preserving evaluation of the source program.

We also describe a prototype implementation and applications of unrestricted intersections and unions: records, operator overloading, and simulating dynamic typing.

Categories and Subject Descriptors F.3.3 [Mathematical Logic and Formal Languages]: Studies of Program Constructs—Type structure

Keywords intersection types

1. Introduction

In type systems, parametric polymorphism is fundamental. It enables generic programming; it supports parametric reasoning about programs. Logically, it corresponds to universal quantification.

Intersection polymorphism (the intersection type $A \land B$) is less well appreciated. It enables ad hoc polymorphism; it supports irregular generic programming. Logically, it roughly corresponds to conjunction.$^1$ Not surprisingly, then, intersection is remarkably versatile.

For both legitimate and historical reasons, intersection types have not been used as widely as parametric polymorphism. One of the legitimate reasons for the slow adoption of intersection types is that no major language has them. A restricted form of intersection, refinement intersection, was realized in two extensions of SML, SML-CIDRE (Davies 2005) and Stardust (Dunfield 2007). These type systems can express properties such as bitwise parity: after refining a type bits of bitstrings with subtypes even (an even number of ones) and odd (an odd number of ones), a bitstring concatenation function can be checked against the type

$$(\text{even} \ast \text{even} \rightarrow \text{even}) \land (\text{odd} \ast \text{odd} \rightarrow \text{even})$$

$$\land (\text{even} \ast \text{odd} \rightarrow \text{odd}) \land (\text{odd} \ast \text{even} \rightarrow \text{odd})$$

which satisfies the refinement restriction: all the intersected types refine a single simple type, bits $\ast$ bits $\rightarrow$ bits.

But these systems were only typecheckers. To compile a program required an ordinary Standard ML compiler. SML-CIDRE was explicitly limited to checking refinements of SML types, without affecting the expressiveness of terms. In contrast, Stardust could typecheck some kinds of programs that used general intersection and union types, but ineffectively: since ordinary SML compilers don’t know about intersection types, such programs could never be run.

Refinement intersections and unions increase the expressiveness of otherwise more-or-less-conventional type systems, allowing more precise properties of programs to be verified through typechecking. The point is to make fewer programs pass the typechecker; for example, a concatenation function that didn’t have the parity property expressed by its type would be rejected. In contrast, unrestricted intersections and unions, in cooperation with a term-level “merge” construct, increase the expressiveness of the term language. For example, given primitive operations $\text{Int.} + : \text{int} \ast \text{int} \rightarrow \text{int}$ and $\text{Real.} + : \text{real} \ast \text{real} \rightarrow \text{real}$, we can easily define an overloaded addition operation by writing a merge:

$$\text{val + } = \text{Int.} + \land \text{Real.} +$$

In our type system, this function + can be checked against the type $[(\text{int} \ast \text{int} \rightarrow \text{int}) \land (\text{real} \ast \text{real} \rightarrow \text{real})]$. In this paper, we consider unrestricted intersection and union types. Central to the approach is a method for elaborating programs with intersection and union types: elaborate intersections into products, and unions into sums. The resulting programs have no intersections and no unions, and can be compiled using conventional means—any SML compiler will do. The above definition of + is elaborated to a pair ($\text{Int.} +$, $\text{Real.} +$): uses of + on ints become first projections of +, while uses on reals become second projections of +.

We present a three-phase design, based on this method, that supports one of our ultimate goals: to develop simpler compilers for full-featured type systems by encoding many features using intersections and unions.

---

$^1$In our setting, this correspondence is strong, as we will see in Sec. 2.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright © 2012 ACM 978-1-4503-1054-3/12/09 ... $10.00
1. An **encoding** phase that straightforwardly rewrites the program, for example, turning a multi-field record type into an intersection of single-field record types, and multi-field records into a “merge” of single-field records.

2. An **elaboration** phase that transforms intersections and unions into products and (disjoint) sums, and intersection and union introductions and eliminations (implicit in the source program) into their appropriate operations: tupling, projection, injection, and case analysis.

3. A **compilation** phase: a conventional compiler with no support for intersections, unions, or the features encoded by phase 1.

**Contributions:** Phase 2 is the main contribution of this paper. Specifically, we will:

- develop elaboration typing rules which, given a source expression $e$ with unrestricted intersections and unions, and a “merging” construct $e_1, e_2$, typecheck and transform the program into an ordinary $\lambda$-calculus term $M$ (with sums and products).
- give a nondeterministic operational semantics ($\rightarrow^*$) for source programs containing merges, in which not all reductions preserve types.
- prove a consistency (simulation) result: ordinary call-by-value evaluation ($\rightarrow^*$) of the elaborated program produces a value corresponding to a value resulting from (type-preserving) reductions of the source program—that is, the diagram in Figure 1 commutes.
- describe an elaborating typechecker that, by implementing the elaboration typing rules, takes programs written in an ML-like language, with unrestricted intersection and union types, and generates Standard ML programs that can be compiled with any SML compiler.

All proofs were checked using the Twelf proof assistant (Pfenning and Schürmann 1999; Twelf 2012) (with the termination checker silenced for a few inductive cases, where the induction measure was nontrivial) and are available on the web (Dunfield and Schürmann 1999; Twelf 2012). For convenience, the names of Twelf source files (\(^2\)) are hyperlinks.

While the idea of compiling intersections to products is not new, this paper is its first full development and practical expression. An essential twist is the source-level merging construct $e_1, e_2$, which embodies several computationally distinct terms, which can be checked against various parts of an intersection type, reminiscent of Forsythe (Reynolds 1996) and (more distantly) the ML-calculus (Castagna et al. 1995). Intersections can still be introduced without this construct; it is required only when no single term can describe the multiple behaviours expressed by the intersection. Remarkably, this merging construct also supports union eliminations with two computationally distinct branches (unlike markers for union elimination in work such as Pierce 1991). As usual, we have no source-level intersection eliminations and no source-level union introductions; elaboration puts all needed projections and injections into the target program.

### Contents:

In Section 2 we give some brief background on intersection types, discuss their introduction and elimination rules, introduce and discuss the merge construct, and compare intersection types to product types. Section 3 gives background on union types, discusses their introduction and elimination rules, and shows how the merge construct is also useful for them. Section 4 has the details of the source language and its (unusual) operational semantics, and describes a non-elaborating type system including subsumption. Section 5 presents the target language and its (entirely standard) typing and operational semantics. Section 6 gives the elaboration typing rules, and proves several key results relating source typing, elaboration typing, the source operational semantics, and the target operational semantics.

Section 7 discusses a major caveat: the approach, at least in its present form, lacks the theoretically and practically important property of coherence, because the meaning of a target program depends on the choice of elaboration typing derivation.

Section 8 shows encodings of type system features into intersections and unions, with examples that are successfully elaborated by our prototype implementation (Section 9). Related work is discussed in Section 10 and Section 11 concludes.

### 2. Intersection Types

What is an intersection type? The simplistic answer is that, supposing that types describe sets of values, $A \land B$ describes the intersection of the sets of values of $A$ and $B$. That is, $v : A \land B$ if $v : A$ and $v : B$.

Less simplistically, the name has been used for substantially different type constructors, though all have a conjunctive flavour. The intersection type in this paper is commutative ($A \land B = B \land A$) and idempotent ($A \land A = A$), following several seminal papers on intersection types (Pottinger 1980; Coppo et al. 1981) and more recent work with refinement intersections (Freeman and Pfenning 1991; Davies and Pfenning 2000; Dunfield and Pfenning 2003). Other lines of research have worked with nonlinear and/or ordered intersections, e.g. [Kfoury and Wells (2004)], which seem less directly applicable to practical type systems (Møller Neegard and Mairson 2004).

For this paper, then: What is a commutative and idempotent intersection type?

One approach to this question is through the Curry-Howard correspondence. Naively, intersection should correspond to logical conjunction—but products correspond to logical conjunction, and intersections are not products, as is evident from comparing the standard introduction and elimination rules for intersection to the (utterly standard) rules for product. (Throughout this paper, $k$ is existentially quantified over $\{1,2\}$; technically, and in the Twelf formulation, we have two rules $\land E_1$ and $\land E_2$, etc.)

$$
\begin{align*}
e : A_1 & \quad e : A_2 \\
\frac{e : A_1 \land A_2}{e : A_1} & \quad \land E_k \\
e : A_1 & \quad e : A_2 \\
\frac{e : A_1 \land A_2 \land E_k}{e : A_2} & \quad \land E_k \\
e : A_1 & \quad e : A_2 \\
\frac{(e_1, e_2) : A_1 \land A_2 \land E_k}{\exists 1} & \quad \land E_k
\end{align*}
$$

\(^2\)For impure call-by-value languages like ML, $\land$ ordinarily needs to be restricted to type a value $v$ for reasons analogous to the value restriction on parametric polymorphism (Davies and Pfenning 2000). Our setting, however, is not ordinary: the technique of elaboration makes the more permissive rule safe, though user-unfriendly. See Section 6.3.

---

**Figure 1.** Elaboration and computation

<table>
<thead>
<tr>
<th>Source language</th>
<th>Target language</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rightarrow, \land, \lor$</td>
<td>$\rightarrow, *, +$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Program</th>
<th>$e : A$</th>
<th>elaboration</th>
<th>$M : T$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nondeterministic evaluation (cbv + merge)</td>
<td>standard evaluation (cbv)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Result</th>
<th>$v : A$</th>
<th>elaboration</th>
<th>$W : T$</th>
</tr>
</thead>
</table>
Here \( \land I \) types a single term \( e \) which inhabits type \( A_1 \) and type \( A_2 \); via Curry-Howard, this means that a single proof term serves as witness to two propositions (the interpretations of \( A_1 \) and \( A_2 \)). On the other hand, in \( \lor I \) two separate terms \( e_1 \) and \( e_2 \) witness the propositions corresponding to \( A_1 \) and \( A_2 \). This difference was suggested by Pottinger (1980), and made concrete when Hindley (1984) showed that intersection (of the form described by Coppo et al. (1981) and Pottinger (1980)) cannot correspond to conjunction because the following type, the intersection of the types of the \( I \) and \( S \) combinators, is uninhabited:

\[
(A \to A) \land ((A \to B \to C) \to (A \to B) \to A \to C)
\]

yet the prospectively corresponding proposition is provable in intuitionistic logic:

\[
(A \lor A) \text{ and} \ ((A \lor B \lor C) \lor (A \lor B) \lor A \lor C) \quad (*)
\]

Hindley notes that every term of type \( A \to A \) is \( \beta \)-equivalent to \( e_1 = \lambda x . x \), and every term of type \( D \) is \( \beta \)-equivalent to \( e_2 = \lambda x . \lambda y . x \cdot x \cdot y \cdot z \cdot s \cdot x \cdot y \cdot z \cdot s \cdot z \), the \( S \) combinator. Any term \( e \) of type \( (A \to A) \land D \) must therefore have two normal forms, \( e_1 \) and \( e_2 \), which is impossible.

But that impossibility holds for the usual \( \land \)-terms. Suppose we add a \textit{merge} construct \( e_{1,n}, e_{2} \) that, quite brazenly, can step to two different things: \( e_{1,n} \to e_{1} \) and \( e_{1,n} \to e_{2} \). Its typing rule chooses one subterm and ignores the other (throughout this paper, the subscript \( k \) ranges over \( \{1, 2\} \)):

\[
e_k : A \quad e_{1,n} : A_1 \quad e_{2} : A_2 \quad \text{merge}_k
\]

In combination with \( \land I \), the \textit{merge} rule allows two distinct implementations \( e_1 \) and \( e_2 \), one for each of the components \( A_1 \) and \( A_2 \) of the intersection:

\[
e_1 : A_1 \quad e_{1,n} : A_1 \quad e_{2} : A_2 \quad e_{1,n} : A_2 \quad \text{merge}_1
\]

Now \( (A \to A) \land D \) is inhabited:

\[
e_{1,n} : A_1 \land A_2
\]

With this construct, the “naive” hope that intersection corresponds to conjunction is realized through elaboration: we can elaborate \( e_{1,n}, e_2 \) to \( (e_1, e_2) \), a term of type \( (A \to A) \land D \), which does correspond to the proposition \((*)\). Inhabitation and provability again correspond—because we have replaced the seemingly mysterious intersections with simple products.

For source expressions, intersection still has several properties that set it apart from product. Unlike product, it has no elimination form. It also lacks an explicit introduction form; \( \land I \) is the only intro rule for \( \land \). While the primary purpose of \( \text{merge}_k \) is to derive the premises of \( \land I \), the \textit{merge} rule makes no mention of intersection (or any other type constructor).

Pottinger (1980) presents intersection \( \land \hat{A} \land B \) as a proposition with some evidence of \( A \) that is also evidence of \( B \)—unlike \( A \land B \), corresponding to \( A \land B \), which has two separate pieces of evidence for \( A \) and for \( B \). In our system, though, \( e_{1,n}, e_2 \) is a single term that provides evidence for \( A \) and \( B \), so it is technically consistent with this view of intersection, but not necessarily consistent in spirit (since \( e_1 \) and \( e_2 \) can be very different from each other).

3. **Union Types**

Having discussed intersection types, we can describe union types as intersections’ dual: if \( v : A_1 \lor A_2 \) then either \( v : A_1 \) or \( v : A_2 \) (perhaps both). This duality shows itself in several ways.

For union \( \lor \), introduction is straightforward, as elimination was straightforward for \( \land \) (again, \( k \) is either \( 1 \) or \( 2 \)):

\[
\Gamma \vdash e : A_k \\
\Gamma \vdash e : A_1 \lor A_2 \lor_k
\]

Coming up with a good elimination rule is trickier. A number of appealing rules are unsound; a sound, yet acceptably strong, rule is

\[
\Gamma \vdash e_0 : A_1 \lor A_2 \\
\Gamma, x_1 : A_1 \vdash \epsilon[x_1] : C \\
\Gamma, x_2 : A_2 \vdash \epsilon[x_2] : C \lor E \\
\Gamma \vdash \epsilon[e_0] : C \lor E
\]

This rule types an expression \( \epsilon[e_0] \)—an evaluation context \( \epsilon \)—with \( e_0 \) in an evaluation position—where \( e_0 \) has the union type \( A_1 \lor A_2 \). During evaluation, \( e_0 \) will be some value \( v_0 \) such that either \( v_0 : A_1 \lor A_2 \). In the former case, the premise \( \Gamma, x_1 : A_1 \vdash \epsilon[x_1] : C \) tells us that substituting \( v_0 \) for \( x_1 \) gives a well-typed expression \( \epsilon[v_0] \). Similarly, the premise \( \Gamma, x_2 : A_2 \vdash \epsilon[x_2] : C \) tells us we can safely substitute \( v_0 \) for \( x_2 \).

The restriction to a single occurrence of \( e_0 \) in an evaluation position is needed for soundness in many settings—generally, in any operational semantics in which \( e_0 \) might step to different expressions. One simple example is a function \( f : (A \to A \to C) \lor (B \to B \to C) \) and expression \( e_0 : A \lor B \), where \( e_0 \) changes the contents pointed to by a reference of type \( (A \lor B) \) ref, before returning the new value. The application \( f e_0 \) would be well-typed by a rule allowing multiple occurrences of \( e_0 \), but unsound: the first \( e_0 \) could evaluate to \( A \) and the second \( e_0 \) to \( B \).

The evaluation context \( \epsilon \) need not be unique, which creates some difficulties for practical typechecking (Dunfield 2011). For further discussion of this rule, see Dunfield and Pfenning (2003).

We saw in Section 2 that, in the usual \( \lambda \)-calculus, \( \land \) does not correspond to conjunction; in particular, no \( \land \)-term behaves like both the \( I \) and \( S \) combinators, so the intersection \( (A \to A) \land D \) (where \( D \) is the type of \( S \)) is uninhabited. In our setting, though, \( (A \to A) \land D \) is inhabited, by the merge of \( I \) and \( S \).

Something similar comes up when eliminating unions. Without the merge construct, certain instances of union types can’t be usefully eliminated. Consider a list whose elements have type \( \text{int} \lor \text{string} \). Introducing those unions to create the list is easy enough: use \( \text{\textit{V1}} \) for the ints and \( \text{\textit{V2}} \) for the strings. Now suppose we want to print a list element \( x : \text{int} \lor \text{string} \), converting the ints to their string representation and leaving the strings alone. To do this, we need a merge: for example, given a function \( g : (\text{\textit{int} \lor \text{string}}) \lor (\text{\textit{string} \lor \text{string}}) \) whose body contains a merge, use rule \( \text{\textit{V1}} \lor \text{\textit{V2}} \) on \( g \) with \( E = g \) \( \text{\textit{V1}} \) \( \text{\textit{V2}} \) \( x \).

Like intersections, unions can be tamed by elaboration. Instead of products, we elaborate products to unions’ dual, sums (tagged unions). Uses of \( \text{\textit{V1}} \lor \text{\textit{V2}} \) become left and right injections into a sum type; uses of \( \text{\textit{V1}} \lor \text{\textit{V2}} \) become ordinary case expressions.

4. **Source Language**

4.1 **Source Syntax**

| Source types | \( A, B, C ::= T | A \to B | A \land B | A \lor B \) |
| Typing contexts | \( \Gamma ::= | \Gamma, x : A \) |
| Source expressions | \( e ::= x | \{ \text{\textit{fix}} \} x . e | e_1 e_2 | \lambda x . e \) |
| Source values | \( v ::= x | \{ \text{\textit{let}} \} x . v | v_{1,2} \) |
| Evaluation contexts | \( \epsilon ::= [1] | \epsilon, e | v \epsilon | \epsilon, e \) |

Figure 2. Syntax of source types, contexts, and expressions
The source language expressions \( e \) are standard, except for the feature central to our approach, the merge \( e_1, e_2 \). The types \( A, B, C \) are a “top” type \( \top \) (which will be elaborated to unit), the usual function space \( A \rightarrow B \), intersection \( A \land B \) and union \( A \lor B \). Values \( v \) are standard, but a merge of values \( v_1, v_2 \) is considered a value, even though it can step! But the step it takes is pure, in the sense that even if we incorporated (say) mutable references, it would not interact with them.

### 4.2 Source Operational Semantics

\[
\begin{align*}
\text{Source expression } e & \text{ steps to } e' \\
\text{step } E & \text{ in } \text{step.elf} \\
\frac{e_1 \leadsto e_1'}{e \leadsto e'} & \text{ step/app1} \\
\frac{e_2 \leadsto e_2'}{\nu_1 e_1 \leadsto \nu_1 e_1'} & \text{ step/app2} \\
\frac{(\lambda x. e) v \leadsto (\lambda y. e) v}{\text{step/beta}} & \\
\frac{\text{fix } x. e \leadsto (\text{fix } x. e)/x e}{\text{step/fix}} & \\
\frac{e_1, e_2 \leadsto e_1, e_2}{\text{step/unmerge left}} & \\
\frac{e_1, e_2 \leadsto e_1, e_2}{\text{step/unmerge right}} & \\
\frac{e_1 \leadsto e_1'}{e_1, e_2 \leadsto e_1, e_2} & \text{ step/merge1} \\
\frac{e_2 \leadsto e_2'}{e_1, e_2 \leadsto e_1, e_2} & \text{ step/merge2} \\
\frac{e \leadsto e_1, e}{\text{step/split}} & \\
\end{align*}
\]

**Figure 3.** Source language operational semantics: call-by-value + merge construct

The source language operational semantics (Figure 3) is standard (call-by-value function application and a fixed point expression) except for the merge construct. This peculiar animal is a descendant of “demonic choice”: by the ‘step/unmerge left’ and ‘step/unmerge right’ rules, \( e_1, e_2 \) can step to either \( e_1 \) or \( e_2 \). Adding to its misbehaviours, it permits stepping within itself (‘step/merge1’ and ‘step/merge2’—note that in ‘step/merge2’, we don’t require \( e_1 \) to be a value). Worst of all, it can appear by spontaneous fission: ‘step/split’ turns any expression \( e \) into a merge of two copies of \( e \).

The merge construct makes our source language operational semantics interesting. It also makes it unrealistic: \( \sim \)-reduction does not preserve types. For type preservation to hold, the operational semantics would need access to the typing derivation. Worse, since the typing rule for merges ignores the unused part of the merge, \( \sim \)-reduction can produce expressions that have no type at all, or are not even closed! The point of the source operational semantics is not to directly model computation; rather, it is a basis for checking that the elaborated program (whose operational semantics is perfectly standard) makes sense. We will show in Section 4.3 that, if the result \( M \) of elaborating \( e \) can step to some \( M' \), then we can step \( e \leadsto^* e' \) where \( e' \) elaborates to \( M' \).

### 4.3 (Source) Subtyping

Suppose we want to pass a function \( f : A \rightarrow C \) to a function \( g : (A \land B) \rightarrow D \). This should be possible, since \( f \) requires only that its argument have type \( A \); in all calls from \( g \) the argument to \( f \) will also have type \( B \), but \( f \) won’t mind. Only the rules discussed so far, however, the application \( g f \) is not well-typed: we can’t get inside the arrow \( (A \land B) \rightarrow C \). For flexibility, we’ll incorporate a subtyping system that can conclude, for example, \( A \rightarrow C \leq (A \land B) \rightarrow C \).

The logic of the subtyping rules (Figure 4, top) is taken straight from Dunfield and Pfenning (2003), so we only briefly give some intuition. Roughly, \( A \leq B \) is sound if every value of type \( A \) can be treated as having type \( B \). Under a subset interpretation, this would mean that \( A \leq B \) is justified if the set of \( A \)-values is a subset of the set of \( B \)-values. For example, the rule \( \forall A \leq B \), if interpreted set-theoretically, says that if \( A \subseteq B_1 \) and \( A \subseteq B_2 \) then \( A \subseteq B_1 \cap B_2 \).

It is easy to show that subtyping is reflexive and transitive; see \[ \text{sub-refl.elf} \] and \[ \text{sub-trans.elf} \] (Building transitivity into the structure of the rules makes it easy to derive an algorithm; an explicit transitivity rule would have premises \( A \leq B \) and \( B \leq C \), which involve an intermediate type \( B \) that does not appear in the conclusion \( A \leq C \).)

Having said all that, the subsequent theoretical development is easier without subtyping. So we will show (Theorem 1) that, given a typing derivation that uses subtyping (through the usual subsumption rule), we can always construct a source expression of the same type that never applies the subsumption rule. This new expression will be the same as the original one, with a few additional coercions. For the example above, we essentially \( \eta \)-expand \( g \) to \( (\lambda x. x) \), which lets us apply \( \land E_1 \) to \( x : A \land B \). Operationally, all the coercions are identities; they serve only to “articulate” the type structure, making subsumption unnecessary.

Note that the coercion in rule \( \forall \land_L \leq \top \to \) is eta-expanded to allow \( \forall E \) to eliminate the union in the type of \( x \); as discussed later, the subexpression of union type must be in evaluation position.

### 4.4 Source Typing

The source typing rules (Figure 4) are either standard or have already been discussed in Sections 2 and 3 except for direct.

The direct rule was introduced and justified in Dunfield and Pfenning (2003, 2004). It is a 1-ary version of \( \forall E \), a sort of cut: a use of the typing \( e_0 : A \) within the derivation of \( E[e_1] : C \) is replaced by a derivations of \( e_0 : A \), along with a derivation of \( E[x] : C \) that assumes \( x : A \). Curiously, in this system of rules, direct is admissible: given \( e_0 : A \), use \( \forall I_1 \) or \( \forall I_2 \) to conclude \( e_0 : A \land A \), then use two copies of the derivation \( x : A \to E[x] : C \) in the premises of \( \forall E \) (a-converting \( x \) as needed). So why include it? Typing using these rules is undecidable; our implementation (Section 9) follows a bidirectional version of them (where typechecking is decidable, given a few annotations, similar to Dunfield and Pfenning (2004)), where direct is not admissible. (A side benefit is that direct and \( \forall E \) are similar enough that it can be helpful to do the direct case of a proof before tackling \( \forall E \).)

**Remark.** Theorem 1 and all subsequent theorems, are proved only for expressions that are closed under the appropriate context, even though \( \text{merge}_0 \) does not explicitly require that the unexamined subexpression be closed; Twelf does not support proofs about objects with unknown variables.

**Theorem 1 (Coercion).** If \( D \) derives \( \Gamma \vdash e : B \) then there exists an \( e' \) such that \( D' \) derives \( \Gamma \vdash e' : B \), where \( D' \) never uses rule sub.

**Proof.** By induction on \( D \). The interesting cases are for sub and \( \forall E \). In the case for sub with \( A \leq B \), we show that when the coercion \( e \mapsto e_{\text{corec}} \) which always has the form \( \lambda x. e_0 \) is applied to an expression of type \( A \), we get an expression of type \( B \). For example, for \( \land L_1 \leq \land E_1 \). This shows that \( e' = (\lambda x. e_0) e \) has type \( B \).

For \( \forall E \), the premises typing \( E[x] \) might “separate”, say if the first includes subsumption (yielding the same \( E[x] \)) and the second doesn’t. Furthermore, inserting coercions could break evaluation positions: given \( E = f \ []) \), replacing \( f \) with an application \( e \mapsto f \)
\[ A \leq B : e \quad \text{Source type } A \text{ is a subtype of source type } B, \]

with coercion \( e \) of type \( \cdot \vdash e : A \rightarrow B \).

\[
\begin{align*}
B_1 \leq A_1 : e & \quad A_2 \leq B_2 : e' \\
\rightarrow & \\
A_1 \\rightarrow A_2 \leq B_1 \rightarrow B_2 : \lambda x. e' (\{ e x \}) \rightarrow \leq & \\
A \leq T : \lambda x. () \rightarrow \leq & \\
A \leq B : e_1 & \quad A \leq B_2 : e_2 \rightarrow \leq & \\
A \leq B_1 \land A_2 \leq B : e \land L \leq & \\
A \leq B : e_1, A_2 \leq B : e_2 \rightarrow \leq & \\
A_1 \lor A_2 \leq B : \lambda x. (\{ e_1, y \}, e_2, y) & \quad x \lor \leq & \\
A \leq B_k : e & \quad A \leq B_1 \lor B_2 : e \lor R_k \leq & \\
\end{align*}
\]

\[ \Gamma \vdash e : A \quad \text{Source expression } e \text{ has source type } A \]

\[
\begin{align*}
\Gamma \vdash e_k : A & \quad \text{merge}_{\Gamma} \\
\Gamma \vdash e_1 : e_2 : A & \quad \text{fix}_{\Gamma} \\
\Gamma \vdash v : T & \quad \text{TI} \\
\Gamma \vdash \lambda x. e : A \rightarrow B & \quad \rightarrow I \\
\Gamma \vdash e_1 : e_2 : B & \quad \rightarrow E \\
\Gamma \vdash e : A_1 & \quad \Gamma \vdash e : A_2 & \quad \land I \\
\Gamma \vdash e : A_1 \lor A_2 & \quad \lor L_k \\
\Gamma \vdash e : A_1 \lor A_2 & \quad \lor E_k \\
\Gamma \vdash e : C & \quad \Gamma \vdash e_0 : C & \quad \lor E \\
\Gamma \vdash e : A & \quad \Gamma \vdash e : B & \quad \lor E \\
\end{align*}
\]

\[ \Gamma \vdash e : A \quad A \leq B \quad \vdash e_{\mathrm{core}} \quad \text{sub} \]

\[ \Gamma \vdash e : B \]

Figure 4. Source type system, with subsumption, non-elaborating

means that \([\cdot]\) is no longer in evaluation position. To handle these issues, let \( e' = (\lambda y. e_1', e_2') e_0 \), where \( e_0 \) comes from applying the induction hypothesis to the derivation of \( \Gamma \vdash e_0 : A_1 \lor A_2 \), and \( e_1' \) and \( e_2' \) come from applying the induction hypothesis to the other two premises. Now \( e_0' \) is in evaluation position, because it follows a \( \lambda \); the merge, typing rule will choose the correct branch.

For details, see \([\text{coerce.elf}]\). We actually encode the typings for \( \text{corec} \) as hypothetical derivations in the subtyping judgment itself \([\text{typeof+sub.elf}]\), making the sub case here trivial.

5. Target Language

Our target language is just the simply-typed call-by-value \( \lambda \)-calculus extended with fixed point expressions, products, and sums.

5.1 Target Syntax

| Target types | \( T \) ::= \text{unit} | T \rightarrow T | T \ast T | T + T |
| Typing contexts | \( G \) ::= \text{cdot} | G, x : T |
| Target terms | \( M, N \) ::= \text{x} | \{ \} | \lambda x. M | M \text{ N} | \text{fix} x. M | (M_1, M_2) | \text{proj}_j. M |
| | | \text{inj}_j. M | \text{case} M \text{ of} \text{inj}_1 x_1 \rightarrow N_1 |
| | | \text{inj}_2 x_2 \rightarrow N_2 |
| Target values | \( W \) ::= \text{x} | \{ \} | \lambda x. M | (W_1, W_2) | \text{inj}_k. W |

Figure 5. Target types and terms

The target types and terms (Figure 5) are completely standard.

5.2 Target Typing

The typing rules for the target language (Figure 5) lack any form of subtyping, and are completely standard.

5.3 Target Operational Semantics

The operational semantics \( M \rightarrow M' \) is, likewise, standard; functions are call-by-value and products are strict. As usual, we write \( M \rightarrow^* M' \) for a sequence of zero or more \( \rightarrow \)s.

Naturally, a type safety result holds:

**Theorem 2** (Target Type Safety). If \( \cdot \vdash M : T \) then either \( M \) is a value, or \( M \rightarrow M' \) and \( \cdot \vdash M' : T \).

**Proof.** By induction on the given derivation, using a few standard lemmas; see \([\text{tm-safety.elf}]\). (The necessary substitution lemma comes for free in Twelf.)

And to calm any doubts about whether \( M \) might step to some other, not necessarily well-typed term:

**Theorem 3** (Determinism of \( \rightarrow \)).
If \( M \rightarrow N_1 \) and \( M \rightarrow N_2 \) then \( N_1 = N_2 \) (up to \text{\( \alpha \)-conversion}.

**Proof.** By simultaneous induction. See \( \text{tm-deterministic} \) in \([\text{tm-safety.elf}]\).

6. Elaboration Typing

We elaborate source expressions \( e \) into target terms \( M \). The source expressions, which include a “merge” construct \( e_1, e_2 \), are typed with intersections and unions, but the result of elaboration is completely standard and can be typed with just unit, \( \rightarrow \), \( \ast \) and \( \rightarrow \).
6.1 Connecting Elaboration and Typing

**Equivalence of elaboration and sourcetyping:** The non-elaborating type assignment system of Figure 4 minus sub, can be read off from the elaboration rules in Figure 9: simply drop the \( \Rightarrow \) part of the judgment. Consequently, given \( e : A \Rightarrow M \) we can always derive \( e : A \):

**Theorem 4.**

If \( \Gamma \vdash e : A \Rightarrow M \) then \( \Gamma \vdash e : A \) (without using rule sub).

**Proof.** By straightforward induction on the given derivation; see `typeof-erase` in `typeof-elab.elf`.

More interestingly, given \( e : A \) we can always elaborate \( e \), so elaboration is just as expressive as typing:

**Theorem 5** (Completeness of Elaboration).

If \( \Gamma \vdash e : A \) (without using rule sub) then \( \Gamma \vdash e : A \Rightarrow M \).

**Proof.** By straightforward induction on the given derivation; see `elab-complete` in `typeof-elab.elf`.

**Elaboration produces well-typed terms:** Any target term \( M \) produced by the elaboration rules has corresponding target type. In the theorem statement, we assume the obvious translation \( \Gamma \), e.g. \( \{x : T \land (T \Rightarrow T) = x ; \{T\} \mid y ; \{T \lor T\} = x : \text{unit} ; y : \text{unit} + \text{unit}\) .

**Theorem 6** (Elaboration Type Soundness).

If \( \Gamma \vdash e : A \Rightarrow M \) then \( \Gamma \vdash M : [A] \).

**Proof.** By induction on the given derivation. For example, the case for direct, which elaborates to an application, applies `typeof/m/arrintro` and `typeof/m/arelim`. Exploiting a bijection between source types and target types, we actually prove \( \Gamma \vdash M : A \), interpreting \( A \) and types in \( \Gamma \) as target types: \( \land \) as \( * \), etc. See `elab-type-soundness.elf`.

---

**Figure 6.** Target type system with functions, products and sums

**Figure 7.** Target language operational semantics:

call-by-value + products + sums

<table>
<thead>
<tr>
<th>( G \vdash M : T )</th>
<th>Target term ( M ) has target type ( T )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{typeofm} \ H \ T \text{ in } [\text{typeofm.elf}] )</td>
<td></td>
</tr>
<tr>
<td>( \text{G, x : T } \vdash M : T \text{ typeofm/ } \text{fix x, M : T } \text{ fix } )</td>
<td></td>
</tr>
<tr>
<td>( \text{G } \vdash M : T \text{ typeofm/ } \text{arrintro} )</td>
<td></td>
</tr>
<tr>
<td>( \text{G } \vdash M : T \rightarrow T' \text{ typeofm/ arelim} )</td>
<td></td>
</tr>
<tr>
<td>( \text{G } \vdash M : (T_1 + T_2) \text{ typeofm/ prodintro} )</td>
<td></td>
</tr>
<tr>
<td>( \text{G } \vdash (\text{inj}<em>k M) : (T_1 + T_2) \text{ typeofm/ sumintro}</em>{k} )</td>
<td></td>
</tr>
<tr>
<td>( \text{G } \vdash M : T_k \text{ typeofm/ sumintro}_{k} )</td>
<td></td>
</tr>
<tr>
<td>( \text{G } \vdash {\text{case } M \text{ of } \text{inj}_1 } \ \text{x}_1 \Rightarrow N_1 \mid \text{inj}_2 \ \text{x}_2 \Rightarrow N_2 \mid T \text{ typeofm/ } \text{sumlim} )</td>
<td></td>
</tr>
</tbody>
</table>

---

**Figure 8.** Type translation

| \( |T| \) | unit |
|---|---|
| \( |A_1 \rightarrow A_2| \) | \( |A_1| \rightarrow |A_2| \) |
| \( |A_1 \land A_2| \) | \( |A_1| \ast |A_2| \) |
| \( |A_1 \lor A_2| \) | \( |A_1| \ast |A_2| \) |

---

6.1 connects elaboration, source typing, and target typing;
6.2 gives lemmas useful for showing that target computations correspond to source computations;
6.3 states and proves that correspondence (consistency, Thm.13);
6.4 summarizes the metatheory through two important corollaries of our various theorems.

Finally, Section 6.5 discusses whether we need a value restriction on \( / \).
6.2 Relating Source Expressions to Target Terms

Elaboration produces a term that corresponds closely to the source expression: a target term is the same as a source expression, except that the intersection- and union-related aspects of the computation become explicit in the target. For instance, intersection elimination via \( \land \), implicit in the source program, becomes the explicit projection \( \text{proj}_1 \). The target term has nearly the same structure as the source; the elaboration rules only insert operations such as \( \text{proj}_1 \), duplicate subterms such as the \( e \) in \( \land \), and omit unused parts of merges.

This gives rise to a relatively simple connection between source expressions and target terms—much simpler than a logical relation, which relates all appropriately-typed terms that have the same extensional behaviour. In fact, stepping in the target preserves elaboration typing, provided we are allowed to step the source expression zero or more times. This consistency result, Theorem 6.3, needs several lemmas.

**Lemma 7.** If \( e \leadsto^* e' \) then \( \mathcal{E}[e] \leadsto^* \mathcal{E}[e'] \).

**Proof.** By induction on the number of steps, using a lemma \((\text{step}-\text{eval-context})\) that \( e \leadsto^* e' \) implies \( \mathcal{E}[e] \leadsto^* \mathcal{E}[e'] \). See \( \text{step-\text{eval-context} in step-\text{eval-context.elf}} \).

Next, we prove inversion properties of unions, intersections and arrows. Roughly, we want to say that if an expression of union type elaborates to an injection \( \text{inj}_k \), it also elaborates to \( M_0 \). For intersections, the property is slightly more complicated: given an expression of intersection type that elaborates to a pair, we can step the expression to get something that elaborates to the components of the pair. Similarly, given an expression of arrow type that elaborates to a \( \lambda \)-abstraction, we can step the expression to a \( \lambda \)-abstraction.

**Lemma 8.** Unions/Injections.

If \( \Gamma \vdash e : A \lor B \vdash \text{inj}_k \rightarrow M_0 \) then \( \Gamma \vdash e : A_k \rightarrow M_0 \).

**Proof.** By induction on the derivation of \( \Gamma \vdash e : C \rightarrow M \). The only possible cases are \( \text{merge}_k \) and \( \lor \text{inj}_k \). See \( \text{elab-\text{intl} and elab-\text{inr} in elab-union.elf} \).

**Lemma 9.** Intersections/Pairs.

If \( \Gamma \vdash e : A \land B \vdash (M_1, M_2) \) then there exist \( e'_1 \) and \( e'_2 \) such that

\[
\Gamma \vdash \text{proj}_1 M_1 \rightarrow M_1, \quad \text{and} \quad \Gamma \vdash \text{proj}_2 M_2 \rightarrow M_2.
\]

**Proof.** By induction on the given derivation; the only possible cases are \( \land \) and merge. See \( \text{elab-\text{sect.elf}} \).

**Lemma 10.** (Arrows/Lambdas).

If \( \Gamma \vdash e : A \rightarrow B \vdash \lambda x.M_0 \) then there exists \( e_0 \) such that \( e \leadsto^* \lambda x.e_0 \) and \( x : A \vdash e_0 : B \vdash M_0 \).

**Proof.** By induction on the given derivation; the only possible cases are \( \land \) and merge. See \( \text{elab-\text{arr.elf}} \).

Our last interesting lemma shows that if an expression \( e \) elaborates to a target value \( W \), we can step \( e \) to some value \( v \) that also elaborates to \( W \).

**Lemma 11.** (Value monotonicity). If \( \Gamma \vdash e : A \rightarrow W \) then \( e \leadsto^* v \) where \( \Gamma \vdash v : A \rightarrow W \).

**Proof.** By induction on the given derivation.

The most interesting case is for \( \land \), where we apply the induction hypothesis to each premise (yielding \( v_1, v_2 \) such that \( e \leadsto^* v_1 \) and \( e \leadsto^* v_2 \)).

In the \( \text{merge}_k \) case on a merge \( e_1, e_2 \), we apply the induction hypothesis to \( e_k \), giving \( e_k \leadsto^* v \). By rule `step/unmerge`, \( e_1, e_2 \leadsto^* e_k \), from which \( e_1, e_2 \leadsto^* v \).

```python
\text{value-\text{mono.elf}}
```

**Lemma 12.** (Substitution).

If \( \Gamma, x : A \vdash e : B \vdash M \) and \( \Gamma \vdash v : A \rightarrow W \) then \( \Gamma \vdash v/x.e : B \vdash [W/x]M \).

**Proof.** By induction on the first derivation. As usual, Twelf gives us this substitution lemma for free.

6.3 Consistency

This theorem is the linchpin: given \( e \) that elaborates to \( M \), we can preserve the elaboration relationship even after stepping \( M \), though we may have to step \( e \) some number of times as well. The expression \( e \) and term \( M \), in general, step at different speeds:

- \( M \) steps while \( e \) doesn’t—for example, if \( M \) is \( \text{proj}_1 \) \( (W_1, W_2) \) and steps to \( W_1 \), there is nothing to do in \( e \) because the projection corresponds to the implicit elimination in rule \( \land \text{E}_1 \);
• e may step more than M—for example, if e is \((v_1, v_2)\) v and M is \((\lambda x. x)\) W, then M \(\beta\) reduces to W, but e must first 'step/unmerge' to the appropriate \(v_k\), yielding \(v_k\) v, and then apply 'step/beta'.

(Note that the converse—if e \(\rightsquigarrow e'\) then M \(\mapsto^{*} M'\)—does not hold: we could pick the wrong half of a merge and get a source expression with no particular relation to M.)

**Theorem 13 (Consistency).**
If \(\vdash e : A \iff M \mapsto M'\) then there exists \(e'\) such that e \(\rightsquigarrow e'\) and \(\vdash e : A \iff M'\).

**Proof.** By induction on the derivation \(D\) of \(\vdash e : A \iff M\). We show several cases here; the full proof is in [consistency.elf].

  - **Case \(\land I\):** Impossible because M cannot step.
  - **Case \(\land D\):**
    \[\vdash e : A_1 \land A_2 \mapsto M_0\]
    \[D : \vdash e : A_1 \land A_2 \mapsto (M_1, M_2)\]
    By inversion, either \(M_1 \mapsto M'_1\) or \(M_2 \mapsto M'_2\). Suppose the former (the latter is similar). By i.h., e \(\rightsquigarrow e'\) and \(\vdash e' : A_1 \mapsto M'_1\).
    For typing, apply merge with premise \(\vdash e' : A_1 \mapsto M'_1\) and with premise \(\vdash e : A_2 \mapsto M_2\).
    Finally, by \(\land I\), we have \(\vdash e' : A_1 \land A_2 \mapsto (M'_1, M_2)\).
  - **Case \(\land E_k\):**
    \[\vdash e : A \land A_2 \mapsto M_0\]
    \[D : \vdash e : A \land e_2 : A \mapsto \text{proj}_k M_0\]
    If \(\text{proj}_k \rightarrow \text{proj}_k M_0\) with \(M_0 \mapsto M_0\), use the i.h. and apply \(\land E_k\).
    \[\text{proj}_k (W_1, W_2) \rightarrow \text{proj}_k M_0 \mapsto W_1, W_2\]
    Using the unrestricted \(\land E\) rule that only types (\(\land\) \(\land\)) \(A\) \(\mapsto A\), and \(\vdash e : A \mapsto M_0\).
    By inversion, either \(e \rightsquigarrow e'\) and \(\vdash e' : A \mapsto M_0\), or \(e \rightsquigarrow e'\) and \(\vdash e : A \mapsto M_0\).
    Finally, by \(\land E_k\), we have \(\vdash e : A \land A_2 \mapsto (M'_1, M_2)\).

**Theorem 14 (Multi-step Consistency).**
If \(\vdash e : A \iff M \mapsto W\) then there exists \(v\) such that \(e \rightsquigarrow^* v\) and \(\vdash v : A \iff W\).

**Proof.** By induction on the derivation of \(M \mapsto^{*} W\). If M is some value \(v\) then, by Lemma 11, \(e\) is some value \(v\). The source expression \(e\) steps to itself in zero steps, so \(e \rightsquigarrow^* v\), and \(\vdash v : A \iff W\) is given (\(e = v\) and \(M = W\)).

Otherwise, we have \(M \mapsto M'\) where \(M' \mapsto^{*} W\). We want to show \(\vdash e' : A \iff M'\), where \(e \rightsquigarrow^* e'\). By Theorem 13 either \(\vdash e : A \mapsto M'\), or \(e \rightsquigarrow e'\) and \(\vdash e' : A \mapsto M'\).

- If \(\vdash e : A \mapsto M'\), let \(e' = e\), so \(\vdash e' : A \mapsto M'\) and \(e \rightsquigarrow^* e'\) in zero steps.
- If \(e \rightsquigarrow e'\) and \(\vdash e' : A \mapsto M'\), we can use the i.h., showing that \(e' \rightsquigarrow^* v\) and \(\vdash v : A \iff W\).

See consistency* in [consistency.elf] \(\Box\)

**6.4 Summing Up**

**Theorem 15 (Static Semantics).**
If \(\vdash e : A\) (using any of the rules in Figure 3) then there exists \(e'\) such that \(\vdash e' : A \iff M \land \vdash M : [A]\).

**Proof.** By Theorems 11(coercion), 3(completeness of elaboration) and 6(enumeration type soundness). [summary.elf] Combined with a run of the target program \((M \mapsto^{*} W)\), they show that elaborated programs are consistent with source programs.

**Theorem 16 (Dynamic Semantics).**
If \(\vdash e : A \iff M \mapsto^{*} W\) then there is a source value \(v\) such that \(e \rightsquigarrow^* v\) and \(\vdash v : A\).

**Proof.** By Theorems 14 (multi-step consistency) and 13.

**6.5 The Value Restriction**

Davies and Pfenning (2009) showed that the then-standard intersection introduction (that is, our \(\land\) I) was unsound in a call-by-value semantics in the presence of effects (specifically, mutable references). Here is an example (modeled on theirs). Assume a base type \(\text{nat} \equiv \{0, 1, 2, \ldots\}\) and a type \(\text{pos} \equiv \{0, 1, 2, \ldots\}\) with values \(0, 1, 2, \ldots\) and a type \(\text{pos} \equiv \{\text{nat} \equiv \{0, 1, 2, \ldots\}\}\). Assume \(\text{pos} \equiv \text{nat}\).

\[\text{let } r = (\text{ref} \ 1) : \text{nat ref} \land (\text{pos ref})\in\]
\[r : \text{ref} \equiv 0;\]
\[(! r) : \text{pos}\]

Using the unrestricted \(\land I\) rule, \(r\) has type \(\text{nat ref} \land (\text{pos ref})\), using \(\land E_1\) yields \(r : \text{nat ref}\), so the write \(r \equiv 0\) is well-typed; using \(\land E_2\) yields \(r : \text{pos ref}\), so the read \(! r\) produces a \(\text{pos}\). In an unelaborated setting, this typing is unsound: \(\text{ref} \equiv 1\) creates a single cell, initially containing \(0\), then overwritten with \(0\), so \(! r \equiv 0\), which does not have type \(\text{pos}\).

Davies and Pfenning proposed, analogously to ML's value restriction on \(\gamma\)-introduction, an \(\land\)-introduction rule that only types values \(v\). This rule is sound with mutable references:

\[
\text{if } \vdash v : A_1 \land v : A_2 \land \vdash (M_1, M_2) \land (A_1, A_2) \land (M_1, M_2) \land (A_1, A_2)
\]

In an elaboration system like ours, however, the problematic example above is sound, because our \(\land I\) elaborates \(\text{ref} \equiv 1\) to two distinct expressions, which create two unaliased cells:
8. Applying Intersections and Unions

8.1 Overloading

A very simple use of unrestricted intersections is to “overload” operations such as multiplication and conversion of data to printable form. SML provides overloading only for a fixed set of built-in operations; it is not possible to write a single square function, as we do in Figure 10. Despite its appearance, \([\text{val} \; \text{sqrt} : \text{int} \rightarrow \text{real} \; | \; \text{real} \rightarrow \text{real}]\) is not a comment but an annotation used to guide our bidirectional typechecker (this syntax, inherited from Stardust, was intended for compatibility with SML compilers, which saw these annotations as comments and ignored them).

In its present form, this idiom is less powerful than type classes (Wadler and Blott 1989). We could extend toString for lists, which would handle lists of integers and lists of reals, but not lists of lists; the version for toString for lists would use the earlier occurrence of toString, defined for integers and reals only. Adding a mechanism for naming a type and then “unioning” it, recursively, is future work.

8.2 Records

Reynolds (1996) developed an encoding of records using intersection types and his version of the merge construct; similar ideas appear in Castagna et al. (1995). Though straightforward, this encoding is more expressive than SML records.

The idea is to add single-field records as a primitive notion, through a type \((\text{fld} : A)\) with introduction form \(\text{fld}(d)\) and the usual eliminations (explicit projection and pattern matching). Once this is done, the multi-field record type \(\{\text{fld1} : A_1, \text{fld2} : A_2\}\) is simply \(\{\text{fld1} : A_1\} \land \{\text{fld2} : A_2\}\), and the corresponding intro form is a merge: \(\text{merge}(\text{fld1} = A_1, \text{fld2} = A_2)\). More standard concrete syntax, such as \(\text{merge}(\text{fld1} = A_1, \text{fld2} = A_2)\), can be handled trivially during parsing.

With subtyping on intersections, we get the desired behaviour of what SML calls “flex records”—records with some fields not listed—with fewer of SML’s limitations. Using this encoding, a function that expects a record with fields x and y can be given any record that has at least those fields, whereas SML only allows one fixed set of fields. For example, the code in Figure 11 is legal in our language but not in SML.

One problem with this approach is that expressions with duplicated field names are accepted. This is part of the larger issue discussed in Section 7.

8.3 Heterogeneous Data

A common argument for dynamic typing over static typing is that heterogeneous data structures are more convenient. For example, dynamic typing makes it very easy to create and manipulate lists containing both integers and strings. The penalty is the loss of compile-time invariant checking. Perhaps the lists should contain integers and strings, but not booleans; such an invariant is not expressible in traditional dynamic typing.

A common rebuttal from advocates of static typing is that it is easy to simulate dynamic typing in static typing. Want a list of integers and strings? Just declare a datatype

```
datatype int_or_string = Int of int | String of string
```

and use int_or_string lists. This guarantees the invariant that the list has only integers and strings, but is unwieldy: each new element must be wrapped in a constructor, and operations on the list elements must unwrap the constructor, even when those operations accept both integers and strings (such as a function of type \(\text{int} \rightarrow \text{string} \land \text{string} \rightarrow \text{string}\)).
(* [ val get_xy : {x:int, y:int} → int * int ] *)
fun get_xy r = (#x(r), #y(r))

(* [ val tupleToString : int * int → string ] *)
fun tupleToString (x, y) = 
  "(" ~ Int.toString x ~ "," ~ Int.toString y ~ ")"
val rec1 = {y = 11, x = 1}
val rec2 = {x = 2, y = 22, extra = 100}
val rec3 = {x = 3, y = 33, other = "a string"}
val _ = print("get_xy rec1 = "
  ~ tupleToString (get_xy rec1) ~ "\n")
val _ = print("get_xy rec2 = "
  ~ tupleToString (get_xy rec2)
  ~ (extra = "
  ~ Int.toString #extra(rec2) ~ ")\n")
val _ = print("get_xy rec3 = "
  ~ tupleToString (get_xy rec3)
  ~ (other = "
  ~ #other(rec3) ~ ")\n")

Output of target program after elaboration:
get_xy rec1 = (1,11)
get_xy rec2 = (2,22) (extra = 100)
get_xy rec3 = (3,33) (other = a string)

Figure 11. Example of flexible multi-field records

datatype 'a list = nil | :: of 'a * 'a list

val toString x = 
  (Int.toString ,
   (fn s ⇒ s : string) ,,
   Real.toString) x

(* [ val hetListToString : dyn list → string ] *)
fun hetListToString xs = case xs of
  nil ⇒ "nil"
| h::t ⇒ (toString h) ~ "::" 
  ~ (hetListToString t)
val _ = print "\n\n"
val _ = print (hetListToString
  [1, 2, "what", 3.14159, 4, "why"])
val _ = print "\n\n"

Output of target program after elaboration:
1::2::what::3.14159::4::why::nil

Figure 12. Example of heterogeneous data

In this situation, our approach provides the compile-time invariant checking of static typing and the transparency of dynamic typing. The type of list elements (if we bother to declare it) is just a union type:

type int_union_string = int ∨ string

Elaboration transforms programs with int_union_string into programs with int_or_string.

Along these lines, we use in Figure 12a type dyn, defined as int ∨ real ∨ string. It would be useful to also allow lists, but the current implementation lacks recursive types of a form that could express "dyn = ... ∨ dyn list".

9. Implementation

Our implementation is faithful to the spirit of the elaboration rules above, but is substantially richer. It is based on Stardust, a typechecker for a subset of core Standard ML with support for inductive datatypes, products, intersections, unions, refinement types and indexed types (Dunfield 2007), extended with support for (first-class) polymorphism (Dunfield 2009). We do not yet support all these features; support for first-class polymorphism looks hardest, since Standard ML compilers cannot even handle higher-rank predicative polymorphism. Elaborating programs that use ML-style prenex polymorphism should work, but we currently lack any proof or even significant testing to back that up.

Our implementation does currently support merges, intersections and unions, a top type, a bottom (empty) type, single-field records and encoded multi-field records (Section 8.2), and inductive datatypes (if their constructors are not of intersection type, though they can take intersections and unions as argument; removing this restriction is a high priority).

9.1 Bidirectional Typechecking

Our implementation uses bidirectional typechecking (Pierce and Turner 2004; Dunfield and Pfenning 2004; Dunfield 2009), an increasingly common technique in advanced type systems; see Dunfield (2009) for references. This technique offers two major benefits over Damas-Milner type inference: it works for many type systems where annotation-free inference is undecidable, and it seems to produce more localized error messages.

Bidirectional typechecking does need more type annotations. However, by following the approach of Dunfield and Pfenning (2004), annotations are never needed except on redexes. The present implementation allows some annotations on redexes to be omitted as well.

The basic idea of bidirectional typechecking is to separate the activity of checking an expression against a known type from the activity of synthesizing a type from the expression itself:

\[ \Gamma \vdash e : A \] e checks against known type A
\[ \Gamma \vdash e : A \] e synthesizes type A

In the checking judgment, \( \Gamma \) and A are inputs to the typing algorithm, which either succeeds or fails. In the synthesis judgment, \( \Gamma \) and e are inputs and A is output (assuming synthesis does not fail).

Syntactically speaking, crafting a bidirectional type system from a type assignment system (like the one in Figure 3) is a matter of taking the colons in the \( \Gamma \vdash e : A \) judgments, and replacing some with "\(:="" and some with "\:⇒". Except for merge, our typing rules can all be found in Dunfield and Pfenning (2004), who argued that introduction rules should check and elimination rules should synthesize. (Parametric polymorphism muddies this picture, but see Dunfield (2009) for an approach used by our implementation.) For functions, this leads to the bidirectional rules

\[ \Gamma, x : A \vdash e : B \] →I
\[ \Gamma, e_1 \equiv e_2 : A \] →E
\[ \Gamma \vdash e_1 \equiv A \] →I
\[ \Gamma \vdash e_1 \equiv e_2 \equiv A \] →E
\[ \Gamma \vdash e_1 \equiv e_2 \equiv A \] →I

The merge rule, however, neither introduces nor eliminates. We implement the obvious checking rule (which, in practice, always tries to check against \( e_1 \) and, if that fails, against \( e_2 \)):

\[ \Gamma \vdash e_1 \equiv e_2 \equiv A \] →I
\[ \Gamma \vdash e_1 \equiv A \] →I
\[ \Gamma \vdash e_2 \equiv A \] →I

Since it can be inconvenient to annotate merges, we also implement synthesis rules, including one that can synthesize an intersection.

\[ \Gamma \vdash e_k : A \] →I
\[ \Gamma \vdash e_1 \equiv A \] →I
\[ \Gamma \vdash e_2 \equiv A \] →I
\[ \Gamma \vdash e_1 \equiv e_2 \equiv A \] →E
\[ \Gamma \vdash e_1 \equiv e_2 \equiv A \] →I
\[ \Gamma \vdash e_1 \equiv A \] →I
\[ \Gamma \vdash e_2 \equiv A \] →I
\[ \Gamma \vdash e_1 \equiv e_2 \equiv A \] →E
\[ \Gamma \vdash e_1 \equiv e_2 \equiv A \] →I
Given a bidirectional typing derivation, it is generally easy to show that a corresponding type assignment exists: replace all "\( \Rightarrow \)" and "\( \leftarrow \)" with "\( \Rightarrow \)" (and erase explicit type annotations from the expression).

9.2 Performance

Intersection typechecking is PSPACE-hard (Reynolds 1996). In practice, we elaborate the examples in Figures 10 [11] and 12 in less than a second, but they are very small. On somewhat larger examples, such as those discussed by Dunfield (2003), the non-elaborating version of Stardust could take minutes, thanks to heavy use of backtracking search (trying \( \setminus E_1 \) then \( \setminus E_2 \), etc.) and the need to check the same expression against different types (\( \setminus I \)) or with different assumptions (\( \setminus E \)). Elaboration doesn’t help with this, but it shouldn’t hurt by more than a constant factor: the shapes of the derivations and the labour of backtracking remain the same.

To scale the approach to larger programs, we will need to consider how to efficiently represent elaborated intersections and unions. Like the theoretical development, the implementation has 2-way intersection and union types, so the type \( A_1 \land A_2 \land A_3 \) is parsed as \( (A_1 \land A_2) \land A_3 \), which becomes \( (A_1 \land A_2) \land A_3 \). A flattened representation \( A_1 \land A_2 \land A_3 \) would be more efficient, except when the program uses values of type \( (A_1 \land A_2) \land A_3 \) where values of type \( A_1 \land A_2 \) are expected; in that case, nesting the product allows the inner pair to be passed directly with no reboxing. Symmetry is also likely to be an issue: passing \( v : A_1 \land A_2 \) where \( v : A_2 \land A_1 \) is expected requires building a new pair. Here, it may be helpful to put the components of intersections into a canonical order.

The foregoing applies to unions as well—introducing a value of a three-way union may require two injections, and so on.

10. Related Work

Intersections were originally developed by Coppo et al. (1981) and Pottinger (1980), among others; Hindley (1992) gives a useful introduction and bibliography. Work on union types began later (MacQueen et al. 1986; Barbanera et al. 1995) is a key paper on type assignment for unions.

Forsythe. In the late 1980s[1], Reynolds invented Forsythe (Reynolds 1997), the first practical programming language based on intersection types. In addition to an unmarked introduction rule like \( \land I \), the Forsythe type system includes rules for typing a construct \( p_1, p_2 \rightarrow _\land \) “a construction for intersecting or ‘merging’ meanings” (Reynolds 1997, p. 24). Roughly analogous to \( e_1, e_2 \), this construct is used to encode a variety of features, but can only be used unambiguously. For instance, a record and a function can be merged, but two functions cannot (actually they can, but the second phrase \( p_2 \) overrides the first). Forsythe does not have union types.

The \( \land \& \)-calculus. Castagna et al. (1995) developed the \( \land \& \)-calculus, which has \&-terms—functions whose body is a merge, and whose type is an intersection of arrows. In their semantics, applying a \&-term to some argument reduces the term to the branch of the merge with the smallest (compatible) domain. Suppose we have a \&-term with two branches, one of type nat \( \rightarrow \) nat and one of type pos \( \rightarrow \) pos. Applying that \&-term to a value of type pos steps to the second branch, because its domain pos is (strictly) a subtype of nat.

Despite the presence of a merge-like construct, their work on the \( \land \& \)-calculus is markedly different from ours: it gives a semantics to programs directly, and uses type information to do so, whereas we elaborate to a standard term language with no runtime type information. In their work, terms have both compile-time types and run-time types (the run-time types become more precise as the computation continues); the semantics of applying a \&-term depends on the run-time type of the argument to choose the branch. The choice of the smallest compatible domain is consistent with notions of inheritance in object-oriented programming, where a class can override the methods of its parent.

Semantic subtyping. Following the \&-calculus, Frisch et al. (2008) investigated a notion of purely semantic subtyping, where the definition of subtyping arises from a model of types, as opposed to the syntactic approach used in our system. They support intersections, unions, function spaces and even complement. Their language includes a dynamic type dispatch which, very roughly, combines a merge with a generalization of our union elimination. Again, the semantics relies on run-time type information.

Pierce’s work. The earliest reference I know for the idea of compiling intersection to product is Pierce (1991b): “a language with intersection types might even provide two different object-code sequences for the two versions of ‘+ [for int and for real]’” (p. 11). Pierce also developed a language with union types, including a term-level construct to explicitly eliminate them (Pierce 1991a).

But this construct is only a marker for where to eliminate the union: it has only one branch, so the same term must typecheck under each assumption. Another difference is that this construct is the only way to eliminate a union type in his system, whereas our \( \land I \) is marker-free. Intersections, also present in his language, have no explicit introduction construct; the introduction rule is like our \( \land I \).

Flow types. Turbak et al. (1997) and Wells et al. (2002) use intersections in a system with flow types. They produce programs with virtual tuples and virtual sums, which correspond to the tuples and sums we produce by elaboration. However, these constructs are internal: nothing in their work corresponds to our explicit intersection and union term constructors, since their system is only intended to capture existing flow properties. They do not compile the virtual constructs into the ordinary ones.

Heterogeneous data and dynamic typing. Several approaches to combining dynamic typing’s transparency and static typing’s guarantees have been investigated. Soft typing (Cartwright and Fagan 1991; Aiken et al. 1994) adds a kind of type inference on top of dynamic typing, but provides no ironclad guarantees. Typed Scheme (Tobin-Hochstadt and Feliesen 2008), developed to retroactively type Scheme programs, has a flow-sensitive type system with union types, directly supporting heterogeneous data in the style of Section 8.3. Unlike soft typing, Typed Scheme guarantees type safety and provides genuine (even first-class) polymorphism, though programmers are expected to provide some annotations.

Type refinements. Restricting intersections and unions to refinements of a single base type simplifies many issues, and is conservative: programs can be checked against refined types, then compiled normally. This approach has been explored for intersections (Free- man and Pfenning 1991; Davies and Pfenning 2000), and for intersections and unions (Dunfield and Pfenning 2003, 2004).

11. Conclusion

We have laid a simple yet powerful foundation for compiling unrestricted intersections and unions; elaboration into a standard functional language. Rather than trying to directly understand the behaviours of source programs, we describe them via their consistency with the target programs.

The most immediate challenge is coherence: While our elaboration approach guarantees type safety of the compiled program, the
meaning of the compiled program depends on the particular elaboration typing derivation used; the meaning of the source program is actually implementation-defined.

One possible solution is to restrict typing of merges so that a merge has type $A$ only if exactly one branch has type $A$. We could also partially revive the value restriction, giving non-values intersection type only if (to a conservative approximation) both components of the intersection are provably disjoint, in the sense that no merge-free expression has both types.

Another challenge is to reconcile, in spirit and form, the unrestricted view of intersections and unions of this paper with the refinement approach. Elaborating a refinement intersection like $(\text{pos} \rightarrow \text{neg}) \land (\text{neg} \rightarrow \text{pos})$ to a pair of functions seems pointless (unless it can somehow facilitate optimizations in the compiler). It will probably be necessary to have “refinement” and “unrestricted” versions of the intersection and union type constructors, at least during elaboration; it may be feasible to hide this distinction at the source level.

Acknowledgments

In 2008, Adam Megacz suggested (after I explained the idea of compiling intersection to product) that one could use an existing ML compiler “as a backend”. The anonymous ICFP reviewers’ suggestions have (I hope) significantly improved the presentation. Finally, I had useful discussions about this work with Yan Chen, Matthew A. Hammer, Scott Kilpatrick, Neelakantan R. Krishnaswami, and Viktor Vafeiadis.

References


