

A Type System for Higher-Order Modules

Derek Dreyer Karl Crary Robert Harper

Carnegie Mellon University

Abstract

We present a type theory for higher-order modules that accounts for many current issues in modular programming languages, including translucency, applicativity, generativity, and modules as first-class values. Our theory harmonizes design elements from various previous work, resulting in a simple, economical, and elegant language. This language is useful as a framework for comparing alternative designs, and is the first language to provide all of these features simultaneously and still support a practical type checking algorithm.

1 Introduction

The design of languages for modular programming is surprisingly delicate and complex. There is a fundamental tension between the desire to separate program components into relatively independent parts and the need to integrate these parts to form a coherent whole. To some extent the design of modularity mechanisms is independent of the underlying language [15], but to a large extent the two are inseparable. For example, languages with polymorphism, generics, or type abstraction require far more complex module mechanisms than those without them.

Much work has been devoted to the design of modular programming languages. Early work on CLU [17] and the Modula family of languages [33, 2] has been particularly influential. Much effort has gone into the design of modular programming mechanisms for the ML family of languages, notably Standard ML [22] and Objective Caml [26]. Numerous extensions and variations of these designs have been considered in the literature [19, 16, 27, 30, 5].

Despite (or perhaps because of) these substantial efforts, the field has remained somewhat fragmented, with no clear unifying theory of modularity having yet emerged. Several competing designs have been proposed, often seemingly at odds with one another. These decisions are as often motivated by pragmatic considerations, such as engineering a useful implementation, as by more fundamental considerations, such as the semantics of type abstraction. The relationship between these design decisions is not completely clear, nor is there a clear account of the trade-offs between them, or whether they can be coherently combined into a single design.

The goal of this paper is to provide a simple, unified formalism for modular programming that consolidates and elucidates much of the work mentioned above. Building on a substantial and growing body of work on type-theoretic accounts of language structure, we propose a type theory for higher-order program modules that harmonizes and enriches these designs and that would be suitable as a foundation for the next generation of modular languages.

1.1 Design Issues

Before describing the main technical features of our language, it is useful to review some of the central issues in the design of module systems for ML. These issues extend to any language of similar expressive power, though some of the trade-offs may be different for different languages.

Controlled Abstraction Modularity is achieved by using signatures (interfaces) to mediate access between program components. The role of a signature is to allow the programmer to “hide” type information selectively. The mechanism for controlling type propagation is *translucency* [9, 12], with transparency and opacity as limiting cases.

Phase Separation ML-like module systems enjoy a *phase separation* property [10] stating that every module is separable into a static part, consisting of type information, and a dynamic part, consisting of executable code. To obtain fully expressive higher-order modules and to support abstraction, it is essential to build this phase separation principle into the definition of type equivalence.

Generativity MacQueen coined the term *generativity* for the creation of “new” types corresponding to run-time instances of an abstraction. For example, we may wish to define a functor `SymbolTable` that, given some parameters, creates a new symbol table. It is natural for the symbol table module to export an abstract type of symbols that are dynamically created by insertion and used for subsequent retrieval. To preclude using the symbols from one symbol table to index another, generativity is essential—each instance of the hash table must yield a “new” symbol type, distinct from all others, even when applied twice to the same parameters.

Separate Compilation One goal of module system design is to support separate compilation [12]. This is achieved by ensuring that all interactions among modules are mediated by interfaces that capture all of the information known to the clients of separately-compiled modules.

Principal Signatures The *principal*, or most expressive, signature for a module captures all that is known about that module during type checking. It may be used as a proxy for that module for purposes of separate compilation. Many type checking algorithms, including the one given in this paper, compute principal signatures for modules.

Modules as First-Class Values Modules in ML are “second-class” in the sense that they cannot be computed as the results of ordinary run-time computation. It can be useful to treat a module as a first-class value that can be stored into

a data structure, or passed as an ordinary function argument or result [9, 23].

Hidden Types Introducing a local, or “hidden”, abstract type within a scope requires that the types of the externally visible components avoid mention of the abstract type. This *avoidance problem* is often a stumbling block for module system design, since in most expressive languages there is no “best” way to avoid a type variable [7, 16].

1.2 A Type System for Modules

The type system proposed here takes into account all of these design issues. It consolidates and harmonizes design elements that were previously seen as disparate into a single framework. For example, rather than regard generativity of abstract types as an alternative to non-generative types, we make both mechanisms available in the language. We support both generative and applicative functors, admit translucent signatures, support separate compilation, and are able to accommodate modules as first-class values [23, 28].

Generality is achieved not by a simple accumulation of features, but rather by isolating a few key mechanisms that, when combined, yield a flexible, expressive, and implementable type system for modules. Specifically, the following mechanisms are crucial.

Singletons Propagation of type sharing is handled by *singleton signatures*, a variant of Aspinall’s and Stone and Harper’s *singleton kinds* [32, 31, 1]. Singletons provide a simple, orthogonal treatment of sharing that captures the full equational theory of types in a higher-order module system with subtyping. No previous module system has provided both abstraction and the full equational theory supported by singletons,¹ and consequently none has provided optimal propagation of type information.

Static Module Equivalence The semantics of singleton signatures is dependent on a (compile-time) notion of equivalence of modules. To ensure that the phase distinction is respected, we define module equivalence to mean “equivalence of static components,” ignoring all run-time aspects.

Subtyping Subtyping is used at the signature level to model “forgetting” type sharing information, which is essential for signature matching. The coercive aspects of signature matching (*e.g.*, dropping of fields and specialization of polymorphic values) are omitted here, since the coercions required are definable in the language.

Determinacy To ensure the proper implementation of abstraction, our type system forbids some modules from being tested for equivalence to any other modules. Modules that are licensed for equality testing are termed *determinate*. Determinacy also provides the license for a module to appear in a signature (via the construction of a singleton signature) or in a type (via projection of one of the module’s type components), as each such appearance implicitly requires the ability to test for equality.

¹Typically the omitted equations are not missed because restrictions to named form or valuability prevent programmers from writing code whose typeability would depend on those equations in the first place [4].

Static and Dynamic Effects The “sealing” [9] of a module with a signature (*aka* “opaque signature ascription”), which is used to create abstract and/or generative types, is deemed by our type system to induce a *pro forma* computational effect. This is backed up by the intuition that generativity involves the generation of new types at run time. For our type system, we set the class of determinate modules to be precisely the pure, effect-free modules. Thus, a module containing any type abstraction via sealing is rendered indeterminate, which is essential to achieving a correct semantics of type abstraction.

The framework of effects allows us to distinguish between, and admit both, generative and non-generative forms of type abstraction. In particular, we isolate two distinct sorts of computational effects, which we call *static* and *dynamic*. Static effects model the hiding of type information at compile time (*i.e.*, non-generative type abstraction), while dynamic effects model the generation of new types at run time (*i.e.*, type generativity). Dynamic effects allow us to encode generativity without resorting to the use of “generative stamps” [22, 19]. Since sealing is the principal means of creating abstract types, we also provide two forms of sealing corresponding to the two kinds of effects: *weak sealing*, which induces a static effect, and *strong sealing*, which induces both a static and a dynamic effect.

Applicative and Generative Functors An *applicative* functor [13] is one that respects static equivalence of its arguments. This models the behavior of functors in Objective Caml. A *generative* functor does not respect equivalence—the abstract types in the result differ on each application. The applicative/generative dichotomy falls out very cleanly from our treatment of abstraction via effects. Specifically, a generative functor is one whose body may contain dynamic effects, corresponding to generative types, while the body of an applicative functor may contain only static effects.

Existential Signatures In a manner similar to Shao [30], our type system is carefully crafted to circumvent the avoidance problem, so that every module enjoys a principal signature. However, this requires imposing restrictions on the programmer. To lift these restrictions, we propose the use of existential signatures to provide principal signatures where none would otherwise exist. We show that these existential signatures are type-theoretically ill-behaved in general, so we restrict their use to a well-behaved setting. In the style of Harper and Stone [11], we propose the use of an elaboration algorithm from an external language that may incur the avoidance problem, into our type system, which does not.

Packaged Modules Modules in our system are “second-class” in the sense that the language of modules is separate from the language of terms. However, following Mitchell *et al.* [23] and Russo [28], we provide a way of packaging a module as a first-class value. In prior work, such packaged modules are typically given an existential type, whose closed-scope elimination construct can make for awkward programming. Instead, our account of type generativity allows us to employ a more natural, open-scope elimination construct, whereby the act of unpackaging a packaged module engenders a dynamic effect.

While these features combine naturally to form a very general language for modular programming, they would be

types	$\tau ::= \text{Typ } M \mid \Pi s:\sigma.\tau \mid \tau_1 \times \tau_2$
terms	$e ::= \text{Val } M \mid \langle e_1, e_2 \rangle \mid \pi_i e \mid e M \mid \text{fix } f(s:\sigma):\tau.e \mid \text{let } s = M \text{ in } (e : \tau)$
signatures	$\sigma ::= 1 \mid \llbracket T \rrbracket \mid \llbracket \tau \rrbracket \mid \Pi s:\sigma_1.\sigma_2 \mid \Pi^{\text{gen}} s:\sigma_1.\sigma_2 \mid \Sigma s:\sigma_1.\sigma_2 \mid \mathfrak{S}(M)$
modules	$M ::= s \mid \langle \rangle \mid \llbracket \tau \rrbracket \mid [e : \tau] \mid \lambda s:\sigma.M \mid M_1 M_2 \mid \langle s = M_1, M_2 \rangle \mid \pi_i M \mid \text{let } s = M_1 \text{ in } (M_2 : \sigma) \mid M :>\sigma \mid M :: \sigma$
contexts	$\Gamma ::= \epsilon \mid \Gamma, s:\sigma$

Figure 1: Syntax

of little use in the absence of a practical implementation strategy. Some previous attempts have encountered difficulties with undecidability [9] or incompleteness of type checking [26]. In contrast, our formalism leads to a practical, implementable programming language.

The rest of this paper is structured as follows: In Section 2 we present our core type system for higher-order modules, including the intuition behind its design and a brief description of the decidable typechecking algorithm. In Section 3 we discuss the programming importance of having both weak and strong forms of sealing. In Section 4 we explain the avoidance problem and how it can be circumvented using an elaboration algorithm. In Section 5 we present a very simple, orthogonal extension of our core system to provide support for packaging modules as first-class values. Finally, in Section 6 we compare our system with related work and in Section 7 we conclude.

2 Technical Development

We begin our technical development by presenting the syntax of our language in Figure 1. Our language consists of four syntactic classes: terms, types, modules, and signatures (which serve as the types of modules). The language does not explicitly include higher-order type constructors or kinds (which ordinarily serve as constructors’ types); in our language the roles of constructors and kinds are subsumed by modules and signatures. Contexts bind module variables (s) to signatures.

As usual, we consider alpha-equivalent expressions to be identical. We write the capture-avoiding substitution of M for s in an expression E as $E[M/s]$.

Types and Terms There are three basic types in our language. The product type ($\tau_1 \times \tau_2$) is standard. The function type, $\Pi s:\sigma.\tau$, is the type of functions that accept a module argument s of signature σ and return a value of type τ (possibly containing s). As usual, if s does not appear free in τ , we write $\Pi s:\sigma.\tau$ as $\sigma \rightarrow \tau$. (This convention is used for the dependent products in the signature class as well.) Finally, when M is a module containing exactly one type (which is to say that M has the signature $\llbracket T \rrbracket$), that type is extracted by $\text{Typ } M$. A full-featured language would support a variety of additional types as well.

The term language contains the natural introduction and elimination constructs for recursive functions and products. In addition, when M is a module containing exactly one value (which is to say that M has the signature $\llbracket \tau \rrbracket$), for

some type τ), that value is extracted by $\text{Val } M$. When f does not appear free in e , we write $\text{fix } f(s:\sigma):\tau.e$ as $\Lambda s:\sigma.e$.

The conventional forms of functions and polymorphic function are built from module functions. Ordinary functions are built using modules containing a single value:

$$\begin{aligned} \tau_1 \rightarrow \tau_2 &\stackrel{\text{def}}{=} \llbracket \tau_1 \rrbracket \rightarrow \tau_2 \\ \lambda x:\tau.e(x) &\stackrel{\text{def}}{=} \Lambda s:\llbracket \tau \rrbracket.e(\text{Val } s) \\ e_1 e_2 &\stackrel{\text{def}}{=} e_1[e_2] \end{aligned}$$

and polymorphic functions are built using modules containing a single type:

$$\begin{aligned} \forall \alpha.\tau(\alpha) &\stackrel{\text{def}}{=} \Pi s:\llbracket T \rrbracket.\tau(\text{Typ } s) \\ \Lambda \alpha.e(\alpha) &\stackrel{\text{def}}{=} \Lambda s:\llbracket T \rrbracket.e(\text{Typ } s) \\ e \tau &\stackrel{\text{def}}{=} e[\tau] \end{aligned}$$

Signatures and Modules There are seven basic signatures in our language. The atomic signature $\llbracket T \rrbracket$ is the type of an atomic module containing a single type, and the atomic signature $\llbracket \tau \rrbracket$ is the type of an atomic module containing a single term. The atomic modules are written $\llbracket \tau \rrbracket$ and $[e : \tau]$, respectively. (We omit the type label “: τ ” from atomic term modules when it is clear from context.) The trivial atomic signature 1 is the type of the trivial atomic module $\langle \rangle$.

The functor signatures $\Pi s:\sigma_1.\sigma_2$ and $\Pi^{\text{gen}} s:\sigma_1.\sigma_2$ express the type of functors that accept an argument of signature σ_1 and return a result of signature σ_2 (possibly containing s). We discuss the difference between Π and Π^{gen} in detail below.

The structure signature $\Sigma s:\sigma_1.\sigma_2$ is the type of a pair of modules where the left-hand component has signature σ_1 and the right-hand component has signature σ_2 , in which s refers to the left-hand component. As usual, when s does not appear free in σ_2 , we write $\Sigma s:\sigma_1.\sigma_2$ as $\sigma_1 \times \sigma_2$.

The singleton signature $\mathfrak{S}(M)$ is used to express type sharing information. It classifies modules that have signature $\llbracket T \rrbracket$ and are statically equivalent to M . Two modules are considered statically equivalent if they are equal modulo term components; that is, type fields must agree but term fields may differ. Singletons at signatures other than $\llbracket T \rrbracket$ are not provided primitively because they can be defined using the basic singleton, as described by Stone and Harper [32]. The definition of $\mathfrak{S}_\sigma(M)$ (the signature containing only modules equal to M at signature σ) is given in Figure 5.

The module syntax contains module variables (s), the atomic modules, and the usual introduction and elimination constructs for Π and Σ signatures, except that Σ modules are introduced by $\langle s = M_1, M_2 \rangle$, in which s stands for M_1 and may appear free in M_2 . (When s does not appear free in M_2 , the “ $s =$ ” is omitted.) No introduction or elimination constructs are provided for singleton signatures. Singletons are introduced and eliminated by rules in the static semantics; if M is judged equivalent to M' in σ , then M belongs to $\mathfrak{S}_\sigma(M')$, and vice versa.

The remaining module constructs are strong sealing, written $M :>\sigma$, and weak sealing, written $M :: \sigma$. When a module is sealed either strongly or weakly, the result is *opaque*. By opaque we mean that no client of the module may depend on any details of the implementation of M other than what is exposed by the signature σ . The distinction between strong and weak sealing is discussed in detail below.

```

signature SIG =
  sig
    type s
    type t = s * int

    structure S : sig
      type u
      val f : u -> s
    end

    val g : t -> S.u
  end
... is compiled as ...

 $\Sigma s: \llbracket T \rrbracket.$ 
 $\Sigma t: \mathfrak{S}(\llbracket \text{Typ } s \times \text{int} \rrbracket).$ 
 $\Sigma S: (\Sigma u: \llbracket T \rrbracket. \Sigma f: \llbracket \text{Typ } u \rightarrow \text{Typ } s \rrbracket. 1).$ 
 $\Sigma g: \llbracket \text{Typ } t \rightarrow \text{Typ}(\pi_1 S) \rrbracket. 1$ 

```

Figure 2: ML Signature Example

```

structure S1 =
  struct
    type s = bool
    type t = bool * int

    structure S = struct
      type u = string
      val f = (fn y:u => true)
    end

    val g = (fn y:t => "hello world")
  end
... is compiled as ...

⟨s = [bool],
  ⟨t = [bool × int],
    ⟨S = ⟨u = [string], ⟨f = [λy: Typ u.true], ⟨⟩⟩⟩,
    ⟨g = [λy: Typ t."hello world"], ⟨⟩⟩⟩⟩⟩

```

Figure 3: ML Structure Example

Although higher-order type constructors do not appear explicitly in our language, they are faithfully represented in our language by unsealed modules containing only type components. For example, the kind $(T \rightarrow T) \rightarrow T$ is represented by the signature $(\llbracket T \rrbracket \rightarrow \llbracket T \rrbracket) \rightarrow \llbracket T \rrbracket$; and the constructor $\lambda\alpha:(T \rightarrow T).(\text{int} \times \alpha \text{int})$ is represented by the module $\lambda s:(\llbracket T \rrbracket \rightarrow \llbracket T \rrbracket).[\text{int} \times \text{Typ}(s[\text{int}])]$.

Examples of how ML-style signatures and structures may be expressed in our language appear in Figures 2 and 3.

2.1 Module Equivalence and Implications

The key issue in the design of our module calculus lies in two related questions: When can modules be compared for equivalence, and, given two comparable modules, when are they deemed equivalent? We say that a module is *determinate* if it is eligible to be compared for equivalence.

Often the literature on module systems has focused on which modules can appear in type projections, rather than

on which modules can be compared for equivalence.² Here we stress equivalence as the primary notion, since the choice of which modules may be compared for equality determines those from which types may be projected.

To see this, suppose that M is an indeterminate module with signature $\llbracket T \rrbracket$. By definition, M cannot be compared for equivalence with any other module. Suppose, however, we are permitted to form the type $\text{Typ } M$. All types are comparable for equality, so $\text{Typ } M$ can be compared for equality with any other type, including $\text{Typ } M'$ (for some other M' with signature $\llbracket T \rrbracket$). Since $\text{Typ } M$ returns the entire content of M , and likewise for M' , this gives us a means by which we may compare M for equivalence with M' , which cannot be permitted. Thus, the type system must permit the projection of type components only from *determinate* modules.

2.1.1 Determinacy

In the literature different accounts of higher-order modules provide different classes of determinate modules. For example, in Harper and Lillibridge’s first-class module system [9], only values are considered determinate. This is necessary in their setting for type soundness because, in the presence of side-effects, non-values could compute different modules containing different types each time they are evaluated. Thus, the type components of a non-value are not well determined (hence the phrase “indeterminate”) and cannot meaningfully be compared for equivalence.

In Leroy’s second-class module calculi [12, 13], determinacy is limited to the syntactic category of paths. In Leroy’s case this was not necessary for type soundness, since modules were never produced by run-time computations, but it served to provide a certain degree of abstraction. In Harper *et al.*’s “phase-distinction” calculus [10], an early higher-order, second-class module system, all modules were deemed to be determinate, but no means of abstraction—such as sealing—was provided.

In our language we wish to admit as large a class of determinate modules as possible. First, we specify determinacy using a semantic condition (formalized by a judgment in the type theory), rather than as a syntactic condition. In addition to being more elegant, this semantic treatment of determinacy turns out to be necessary for a correct account of generativity. Second, we deem all modules to be determinate unless indeterminacy is imposed by programmer-specified abstraction and/or generativity.

Abstraction via Sealing The principal means for defining abstract types is *sealing*, written $M :> \sigma$. Sealing M with σ prevents any client of M from depending on the identities of any type components specified opaquely—*i.e.*, with signature $\llbracket T \rrbracket$ —inside σ . When considered from the point of view of module equivalence, this opacity means that a sealed module should not be considered determinate.

To see this, suppose that $M = ([\text{int}] :> \llbracket T \rrbracket)$ is determinate. Presumably, M will not be deemed equivalent to $[\text{bool}] :> \llbracket T \rrbracket$. However, since module equivalence is reflexive, M will be deemed equivalent to itself. For the type system to judge two opaque modules— M and M —as equivalent just because they have the same underlying implementation is a clear violation of the principle of type abstraction. Thus,

²For example, Harper *et al.* [10] included an explicit notion of module equivalence, but Harper and Lillibridge [9] and Leroy [12, 13] did not.

it is imperative that a module containing any form of sealing be judged indeterminate.

Effects and Purity The irreflexivity of opaque module expressions is strongly reminiscent of the irreflexivity of expressions in languages with effects. In a first-class module system such as Harper and Lillibridge’s [9], an effectful module expression must be ruled indeterminate, since it could yield a distinct type each time it is evaluated. In a second-class module system no such behavior is, in fact, possible, but it is useful to regard a sealed module as “hiding” an arbitrary computational effect.

This brings about the familiar notions of *pure* (effect-free) and *impure* (effectful) expressions, namely by considering sealing to induce an effect.³ With this in mind, we may define the set of determinate modules provided by our type system: a module is determinate if and only if it is pure. This follows our intuition, as the meaning of an effectful expression is not “well-determined,” and it is necessary to provide the desired abstraction.

Applicative vs. Generative Functors One of the key issues in module system design revolves around the interaction of abstract types and functors. Functors in Standard ML are *generative* in the sense that they generate distinct abstract types at each application. Functors in Objective Caml, however, are non-generative, or *applicative*. When applied to the same argument twice, an O’Caml functor yields results with equivalent abstract type components.

In order to model both kinds of functors in our system, we distinguish between two different forms of sealing—*strong*, written $M :> \sigma$ as before, and *weak*, written $M :: \sigma$. Both are opaque, as described above, but only strong sealing is *generative*. By “generative,” we mean that each time the module expression $M :> \sigma$ is evaluated at run time, it “generates” abstract types distinct to that dynamic instance. In particular, a functor whose body is strongly sealed should generate distinct abstract types every time it is applied, as each application constitutes a separate dynamic instance of the body.

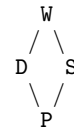
Weak sealing, on the other hand, allows one to write a functor whose body is sealed opaquely, but which does not generate fresh abstract types at each application. In other words, while $M :> \sigma$ is viewed as generating new type components at run time, $M :: \sigma$ is viewed as generating new type components at *compile time*. As a result, it is reasonable to distinguish between the sort of effects induced by weak and strong sealing. Specifically, opacity corresponds to a *static* effect, while generativity corresponds to a *dynamic* effect. Thus, weak sealing induces a static effect only, while strong sealing induces both a static and a dynamic effect. We discuss the importance of supporting *both* weak and strong sealing in Section 3.

Static vs. Dynamic Effects In our type system, as usual, dynamic effects are suspended when they appear within a lambda. Thus, a lambda is always dynamically pure (free of dynamic effects). When such a lambda is applied, whatever dynamic effects were suspended within it are released, and so a functor application may or may not have dynamic

effects. In order that our type system can track dynamic effects, we use signatures to distinguish between functors that may or may not induce dynamic effects when applied. An applicative functor—*i.e.*, one with a Π signature—always has a dynamically pure body, and thus its application (to a dynamically pure argument) is dynamically pure. However, a generative functor—*i.e.*, one with a Π^{gen} signature—may have a dynamically impure body, and thus may generate dynamic effects when applied. One consequence is that the application of a generative functor must *not* be deemed determinate (even if the functor and its argument both are).

In contrast, static effects, being generated at compile time, may be resolved under a lambda. More generally, no construct can suspend a static effect. Therefore, any module containing any strong or weak sealing is statically impure and therefore indeterminate. Nevertheless, a dynamically pure, statically impure module enjoys some privileges not enjoyed by modules in general. In particular, it can appear within the body of an applicative functor. Thus opaque sub-structures (such as ML-style datatypes [11]) can appear within applicative functors. Conversely, a statically pure, dynamically impure module may only appear in the body of a generative functor, but that functor then captures the dynamic effects, resulting in a determinate functor.

Formalization The rules of our type system follow from these definitions. We write our typing judgment $\Gamma \vdash_{\kappa} M : \sigma$, where κ indicates M ’s purity. The classifier κ is drawn from the lattice:



where P indicates that M is pure (and hence determinate), D indicates dynamic purity, S indicates static purity, and W indicates well-formedness only (no purity information). Hence, $\Gamma \vdash_P M : \sigma$ is our determinacy judgment. It will prove to be convenient in our typing rules to exploit the ordering (written \sqsubseteq), meets (\sqcap), and joins (\sqcup) of this lattice, where P is taken as the bottom and W is taken as the top. We also sometimes find it convenient to write functor signatures as $\Pi^{\delta} s : \sigma_1 . \sigma_2$, where $\delta \in \{\epsilon, \text{gen}\}$ and take $\epsilon \sqsubseteq \text{gen}$.

Some key rules are summarized in Figure 4: Pure modules are dynamically pure and statically pure, and each of those are at least well-formed (rule 1). Strongly sealed modules are neither statically nor dynamically pure (2); weakly sealed modules are not statically pure, but are dynamically pure if their body is (3). Applicative functors must have dynamically pure bodies (5); generative functors have no restriction (6). Applicative functors may be used as generative ones (7). Variables are pure (4), and lambdas are dynamically pure (5 and 6). The application of an applicative functor is as pure as the functor itself (8), but the application of a generative functor is at best statically pure (9). Finally, the purity of a module is preserved by signature subsumption (12). The complete set of typing rules is given in Appendix A.

The rules for functor application (rules 8 and 9) require that the functor argument be determinate. This is because the functor argument is substituted into the functor’s codomain to produce the result signature, and the substitution of indeterminate modules for variables (which are

³Since we are working with a second-class module system, this is the only source of impurity in the language. Effectful expressions in the core language do not introduce an impurity in our sense.

$$\begin{array}{c}
\frac{\Gamma \vdash_{\kappa} M : \sigma \quad \kappa \sqsubseteq \kappa'}{\Gamma \vdash_{\kappa'} M : \sigma} \quad (1) \qquad \frac{\Gamma \vdash_{\kappa} M : \sigma}{\Gamma \vdash_{\mathbb{W}} (M : > \sigma) : \sigma} \quad (2) \qquad \frac{\Gamma \vdash_{\kappa} M : \sigma}{\Gamma \vdash_{\kappa \sqcup \text{D}} (M :: \sigma) : \sigma} \quad (3) \qquad \frac{\Gamma \vdash \text{ok}}{\Gamma \vdash_{\text{P}} s : \Gamma(s)} \quad (4) \\
\\
\frac{\Gamma, s : \sigma_1 \vdash_{\kappa} M : \sigma_2 \quad \kappa \sqsubseteq \text{D}}{\Gamma \vdash_{\kappa} \lambda s : \sigma_1. M : \Pi s : \sigma_1. \sigma_2} \quad (5) \qquad \frac{\Gamma, s : \sigma_1 \vdash_{\kappa} M : \sigma_2}{\Gamma \vdash_{\kappa \sqcap \text{D}} \lambda s : \sigma_1. M : \Pi^{\text{gen}} s : \sigma_1. \sigma_2} \quad (6) \qquad \frac{\Gamma, s : \sigma_1 \vdash \sigma_2 \text{ sig}}{\Gamma \vdash \Pi s : \sigma_1. \sigma_2 \leq \Pi^{\text{gen}} s : \sigma_1. \sigma_2} \quad (7) \\
\\
\frac{\Gamma \vdash_{\kappa} M_1 : \Pi s : \sigma_1. \sigma_2 \quad \Gamma \vdash_{\text{P}} M_2 : \sigma_1}{\Gamma \vdash_{\kappa} M_1 M_2 : \sigma_2 [M_2 / s]} \quad (8) \qquad \frac{\Gamma \vdash_{\kappa} M_1 : \Pi^{\text{gen}} s : \sigma_1. \sigma_2 \quad \Gamma \vdash_{\text{P}} M_2 : \sigma_1}{\Gamma \vdash_{\kappa \sqcup \text{S}} M_1 M_2 : \sigma_2 [M_2 / s]} \quad (9) \\
\\
\frac{\Gamma \vdash_{\kappa} M : \Sigma s : \sigma_1. \sigma_2}{\Gamma \vdash_{\kappa} \pi_1 M : \sigma_1} \quad (10) \qquad \frac{\Gamma \vdash_{\text{P}} M : \Sigma s : \sigma_1. \sigma_2}{\Gamma \vdash_{\text{P}} \pi_2 M : \sigma_2 [\pi_1 M / s]} \quad (11) \qquad \frac{\Gamma \vdash_{\kappa} M : \sigma \quad \Gamma \vdash \sigma \leq \sigma'}{\Gamma \vdash_{\kappa} M : \sigma'} \quad (12)
\end{array}$$

Figure 4: Key Typing Rules

always determinate) can turn well-formed signatures into ill-formed ones (for example, $[\text{Typ } s]$ is ill-formed when an indeterminate module is substituted for s). (The alternative rule proposed by Harper and Lillibridge [9] resolves this issue, but induces the avoidance problem, as we discuss in Section 4.) Therefore, when a functor is to be applied to an indeterminate argument, that argument must first be bound to a variable, thereby making it determinate. Similarly, projection of the second component of a pair is restricted to determinate pairs (rule 11), but no such restriction need be made for projection of the first component (rule 10).

2.1.2 Static Equivalence

The second key design issue is, when are determinate modules deemed equivalent? If a determinate module has signature $[[T]]$, it is possible to extract types from it. Type checking depends essentially on the matter of which types are equal, so we must consider when $\text{Typ } M$ is equal to $\text{Typ } M'$.

The simplest answer to this question would be that $\text{Typ } M = \text{Typ } M'$ exactly when the modules M and M' are equal. This simple answer is naive because we cannot in general determine when two modules are equal. Suppose $F : [[\text{int}]] \rightarrow \sigma$ and $e, e' : \text{int}$. Then $F[e] = F[e']$ if and only if $e = e'$, but the latter equality is undecidable in general. To preserve the phase distinction [10] between compile-time and run-time computation, we must ensure that type checking never requires the evaluation of program terms.

To resolve this problem, observe that our language provides no means by which a type component of a module can depend on a term component. (This is not happenstance, but the result of careful design. We will see in Section 5 that the matter is more subtle than it appears.) Consequently, we may ignore general equality and restrict our attention to the *static* equivalence of modules. Two modules are deemed to be equivalent if they agree on all type components.⁴ In the sequel, we will sometimes refer to “static equivalence” to emphasize the static nature of our equivalence.

⁴The phase distinction calculus of Harper, *et al.* [10] includes “non-standard” equality rules for phase-splitting modules M into structures $\langle M_{\text{stat}}, M_{\text{dyn}} \rangle$ consisting of a static component M_{stat} and a dynamic component M_{dyn} . Our static equivalence $M \cong M'$ amounts to saying $M_{\text{stat}} = M'_{\text{stat}}$ in their system. However, we do not identify functors with structures, as they do.

We write our module equivalence judgment as $\Gamma \vdash M \cong M' : \sigma$. The rules for static equivalence of atomic modules are the expected ones. Atomic type components must be equal, but atomic term components need not be:

$$\frac{\Gamma \vdash \tau \equiv \tau'}{\Gamma \vdash [\tau] \cong [\tau'] : [[T]]} \qquad \frac{\Gamma \vdash_{\text{P}} M : [[\tau]] \quad \Gamma \vdash_{\text{P}} M' : [[\tau']]}{\Gamma \vdash M \cong M' : [[\tau]]}$$

Since the generative production of new types in a generative functor is notionally a dynamic operation, generative functors have no static components to compare. Thus, determinate generative functors are always statically equivalent, just as atomic term modules are:

$$\frac{\Gamma \vdash_{\text{P}} M : \Pi^{\text{gen}} s : \sigma_1. \sigma_2 \quad \Gamma \vdash_{\text{P}} M' : \Pi^{\text{gen}} s : \sigma_1. \sigma_2}{\Gamma \vdash M \cong M' : \Pi^{\text{gen}} s : \sigma_1. \sigma_2}$$

The complete set of equivalence rules is given in Appendix A.

As an aside, the notion of static equivalence refutes the misconception that first-class modules are more general than second-class modules. In fact, first- and second-class modules are incomparable. First-class modules have the obvious advantage that they are first-class. However, since the type components of a first-class module can depend on run-time computations, it is impossible to get by with static module equivalence and one must use dynamic equivalence instead (in other words, one cannot phase-split modules as in Harper *et al.* [10]). Consequently, first-class modules cannot propagate as much type information as second-class modules can.

2.2 Singleton Signatures

Type sharing information is expressed in our language using singleton signatures [32], a derivative of translucent sums [9, 12, 16]. (An illustration of the use of singleton signatures to express type sharing appears in Figure 2.) The type system allows the deduction of equivalences from membership in singleton signatures, and vice versa, and also allows the forgetting of singleton information using the sub-signature relation:

$$\frac{\Gamma \vdash_{\text{P}} M : \mathfrak{S}_{\sigma}(M') \quad \Gamma \vdash_{\text{P}} M' : \sigma}{\Gamma \vdash M \cong M' : \sigma} \qquad \frac{\Gamma \vdash M \cong M' : \sigma}{\Gamma \vdash_{\text{P}} M : \mathfrak{S}_{\sigma}(M')}$$

$$\frac{\Gamma \vdash_p M : \sigma}{\Gamma \vdash \mathfrak{S}_\sigma(M) \leq \sigma} \quad \frac{\Gamma \vdash M \cong M' : \sigma}{\Gamma \vdash \mathfrak{S}_\sigma(M) \leq \mathfrak{S}_\sigma(M')}$$

When $\sigma = \llbracket T \rrbracket$, these deductions follow using primitive rules of the type system (since $\mathfrak{S}_{\llbracket T \rrbracket}(M) = \mathfrak{S}(M)$ is primitive). At other signatures, they follow from the definitions given in Figure 5.

Beyond expressing sharing, singletons are useful for “self-ification” [9]. For instance, if s is a variable bound with the signature $\llbracket T \rrbracket$, s can be given the fully transparent signature $\mathfrak{S}(s)$. This fact is essential to the existence of principal signatures in our type checking algorithm. Note that since singleton signatures express static equivalence information, the formation of singleton signatures is restricted to determinate modules. Thus, only determinate modules can be selfified (as in Harper and Lillibridge [9] and Leroy [12]).

Singleton signatures complicate equivalence checking, since equivalence can depend on context. For example, $\lambda s:\llbracket T \rrbracket.\text{[int]}$ and $\lambda s:\llbracket T \rrbracket.s$ are obviously inequivalent at signature $\llbracket T \rrbracket \rightarrow \llbracket T \rrbracket$. However, using subsignatures, they can also be given the signature $\mathfrak{S}(\text{[int]}) \rightarrow \llbracket T \rrbracket$ and at that signature they *are* equivalent, since they return the same result when given the only permissible argument, [int] .

As this example illustrates, the context sensitivity of equivalence provides more type equalities than would hold if equivalence were strictly context insensitive, thereby allowing the propagation of additional type information. For example, if $F : (\mathfrak{S}(\text{[int]}) \rightarrow \llbracket T \rrbracket) \rightarrow \llbracket T \rrbracket$, then the types $\text{Typ}(F(\lambda s:\llbracket T \rrbracket.\text{[int]}))$ and $\text{Typ}(F(\lambda s:\llbracket T \rrbracket.s))$ are equal, which could not be the case under a context-insensitive regime.

A subtle technical point arises in the use of the higher-order singletons defined in Figure 5. Suppose $F : \llbracket T \rrbracket \rightarrow \llbracket T \rrbracket$. Then $\mathfrak{S}_{\llbracket T \rrbracket \rightarrow \llbracket T \rrbracket}(F) = \Pi s:\llbracket T \rrbracket.\mathfrak{S}(Fs)$, which intuitively contains the modules equivalent to F : those that take members of F 's domain and return the same thing that F does. Formally speaking, however, the canonical member of this signature is not F but its eta-expansion $\lambda s:\llbracket T \rrbracket.Fs$. In fact, it is not obvious that F belongs to $\mathfrak{S}_{\llbracket T \rrbracket \rightarrow \llbracket T \rrbracket}(F)$.

To ensure that F belongs to its singleton signature, our type system (following Stone and Harper [32]) includes the extensional typing rule:

$$\frac{\Gamma \vdash_p M : \Pi s:\sigma_1.\sigma_2' \quad \Gamma, s:\sigma_1 \vdash_p M s : \sigma_2}{\Gamma \vdash_p M : \Pi s:\sigma_1.\sigma_2}$$

Using this rule, F belongs to $\Pi s:\llbracket T \rrbracket.\mathfrak{S}(Fs)$ because it is a function and because Fs belongs to $\mathfrak{S}(Fs)$. A similar extensional typing rule is provided for products.

It is possible that the need for these rules could be avoided by making higher-order singletons primitive, but we have not explored the metatheoretic implications of such a change.

2.3 Type Checking

Our type system enjoys a sound, complete, and effective type checking algorithm. Our algorithm comes in three main parts: first, an algorithm for synthesizing the principal (*i.e.*, minimal) signature of a module; second, an algorithm for checking subsignature relationships; and third, an algorithm for deciding equivalence of modules and of types.

Module typechecking then proceeds in the usual manner, by synthesizing the principal signature of a module and then checking that it is a subsignature of the intended signature. The signature synthesis algorithm is given in Appendix B, and its correctness theorems are stated below. The main

$$\begin{aligned} \mathfrak{S}_{\llbracket T \rrbracket}(M) &\stackrel{\text{def}}{=} \mathfrak{S}(M) \\ \mathfrak{S}_{\llbracket \tau \rrbracket}(M) &\stackrel{\text{def}}{=} \llbracket \tau \rrbracket \\ \mathfrak{S}_1(M) &\stackrel{\text{def}}{=} 1 \\ \mathfrak{S}_{\Pi s:\sigma_1.\sigma_2}(M) &\stackrel{\text{def}}{=} \Pi s:\sigma_1.\mathfrak{S}_{\sigma_2}(Ms) \\ \mathfrak{S}_{\Pi^{\text{gen}} s:\sigma_1.\sigma_2}(M) &\stackrel{\text{def}}{=} \Pi^{\text{gen}} s:\sigma_1.\sigma_2 \\ \mathfrak{S}_{\Sigma s:\sigma_1.\sigma_2}(M) &\stackrel{\text{def}}{=} \mathfrak{S}_{\sigma_1}(\pi_1 M) \times \\ &\quad \mathfrak{S}_{\sigma_2[\pi_1 M/s]}(\pi_2 M) \\ \mathfrak{S}_{\mathfrak{S}(M')}(M) &\stackrel{\text{def}}{=} \mathfrak{S}(M) \end{aligned}$$

Figure 5: Singletons at Higher Signatures

judgment of signature synthesis is $\Gamma \vdash_\kappa M \Rightarrow \sigma$, which states that M 's principal signature is σ and M 's purity is inferred to be κ .

Subsignature checking is syntax-directed and easy to do, given an algorithm for checking module equivalence; module equivalence arises when two singleton signatures are compared for the subsignature relation. The equivalence algorithm is closely based on Stone and Harper's algorithm [32] for type constructor equivalence in the presence of singleton kinds. Space considerations preclude further discussion of this algorithm here. Full details of all these algorithms and proofs appear in the companion technical report [6].

Theorem 2.1 (Soundness) *If $\Gamma \vdash_\kappa M \Rightarrow \sigma$ then $\Gamma \vdash_\kappa M : \sigma$.*

Theorem 2.2 (Completeness) *If $\Gamma \vdash_\kappa M : \sigma$ then $\Gamma \vdash_{\kappa'} M \Rightarrow \sigma'$ and $\Gamma \vdash \sigma' \leq \sigma$ and $\kappa' \sqsubseteq \kappa$.*

Note that since the synthesis algorithm is deterministic, it follows from Theorem 2.2 that principal signatures exist. Finally, since our synthesis algorithm, for convenience, is presented in terms of inference rules, we require one more result stating that it really is an algorithm:

Theorem 2.3 (Effectiveness) *For any Γ and M , it is decidable whether there exist σ and κ such that $\Gamma \vdash_\kappa M \Rightarrow \sigma$.*

3 Strong vs. Weak Sealing

Generativity is essential for providing the necessary degree of abstraction in the presence of effects. When a module has side-effects, such as the allocation of storage, abstraction may demand that types be generated in correspondence to storage allocation, in order to ensure that elements of those types relate to the local store and not the store of another instance.

Consider, for example, the symbol table example given in Figure 6. A symbol table contains an abstract type `symbol`, operations for interconverting symbols and strings, and an equality test (presumably faster than that available for strings). The implementation creates an internal hash table and defines symbols to be indices into that internal table.

The intention of this implementation is that the `Fail` exception never be raised. However, this depends on the generativity of the `symbol` type. If another instance, `SymbolTable2`, is created, and the types

```

signature SYMBOL_TABLE =
  sig
    type symbol
    val string_to_symbol : string -> symbol
    val symbol_to_string : symbol -> string
    val eq : symbol * symbol -> bool
  end

functor SymbolTableFun () :> SYMBOL_TABLE =
  struct
    type symbol = int

    val table : string array =
      (* allocate internal hash table *)
      Array.array (initial_size, NONE)

    fun string_to_symbol x =
      (* lookup (or insert) x *) ...

    fun symbol_to_string n =
      (case Array.sub (table, n) of
        SOME x => x
       | NONE => raise (Fail "bad symbol"))

    fun eq (n1, n2) = (n1 = n2)
  end

structure SymbolTable = SymbolTableFun ()

```

Figure 6: Strong Sealing Example

`SymbolTable.symbol` and `SymbolTable2.symbol` are considered equal, then `SymbolTable` could be asked to interpret indices into `SymbolTable2`'s table, thereby causing failure. Thus, it is essential that `SymbolTable.symbol` and `SymbolTable2.symbol` be considered unequal.

The symbol table example demonstrates the importance of strong sealing for encoding generative abstract types in stateful modules. Generativity is not necessary, however, for purely functional modules. Leroy [13] gives several examples of such modules as motivation for the adoption of applicative functors. For instance, one may wish to implement persistent sets using ordered lists. Figure 7 exhibits a purely functional `SetFun` functor, which is parameterized over an ordered element type, and whose implementation of the abstract `set` type is sealed. When `SetFun` is instantiated multiple times—*e.g.*, in different client modules—with the same element type, it is useful for the resulting abstract `set` types to be seen as interchangeable.

In our system, `SetFun` is made applicative, but still opaque, by *weakly* sealing its body. Specifically, `IntSet1.set` and `IntSet2.set` are both equivalent to `SetFun(IntOrd).set`. This type is well-formed because `SetFun` has an applicative functor signature, and `SetFun` and `IntOrd`, being variables, are both determinate. Recall that a functor containing weak sealing is indeterminate and must be bound to a variable before it can be used applicatively.

The astute reader may notice that weak sealing is not truly necessary in the `SetFun` example. In fact, one can achieve the same effect as the code in Figure 7 by leaving the body of the functor unsealed and (strongly) sealing

```

signature ORD =
  sig
    type elem
    val compare : elem * elem -> order
  end

signature SET = (* persistent sets *)
  sig
    type elem
    type set
    val empty : set
    val insert : elem * set -> set
    ...
  end

functor SetFun (Elem : ORD)
  :: SET where type elem = Elem.elem =
  struct
    type elem = Elem.elem
    type set = elem list
    ...
  end

structure IntOrd = struct
  type elem = int
  val compare = Int.compare
end

structure IntSet1 = SetFun(IntOrd)
structure IntSet2 = SetFun(IntOrd)

```

Figure 7: Weak Sealing Example

the functor itself with an applicative functor signature before binding it to `SetFun`. This is the technique employed by Shao [30] for encoding applicative functors, as his system lacks an analogue of weak sealing. A failing of this approach is that it only works if the functor body is fully transparent—in the absence of weak sealing, any opaque substructures would have to be strongly sealed, preventing the functor from being given an applicative signature.

The best examples of the need for opaque substructures in applicative functors are provided by the interpretation of ML datatypes as abstract types [11]. In both Standard ML and Caml datatypes are *opaque* in the sense that their representation as recursive sum types is not exposed. Standard ML and Caml differ, however, on whether datatypes are *generative*. In the presence of applicative functors (which are absent from Standard ML) there is excellent reason for datatypes *not* to be generative—namely, that a generative interpretation would prevent datatypes from appearing in the bodies of applicative functors. This would severely diminish the utility of applicative functors, particularly since in ML recursive types are provided only through the datatype mechanism. For example, an implementation of `SetFun` with splay trees, using a `datatype` declaration to define the tree type, would require the use of weak sealing.

For these reasons, strong sealing is no substitute for weak sealing. Neither is weak sealing a substitute for strong. As Leroy [13] observed, in functor-free code, generativity can be simulated by what we call weak sealing. (This can be seen in our framework by observing that dynamic purity provides no extra privileges in the absence of functors.) With func-

tors, however, strong sealing is necessary to provide true generativity. Nevertheless, it is worth noting that strong sealing is definable in terms of other constructs in our language, while weak sealing is not. In particular, we can define strong sealing, using a combination of weak sealing and generative functor application, as follows:

$$M \text{ :> } \sigma \stackrel{\text{def}}{=} ((\lambda _ . 1.M) \text{ : : } (1 \xrightarrow{\text{gen}} \sigma)) \langle \rangle$$

The existence of this encoding does not diminish the importance of strong sealing, which we have made primitive in our language regardless.

4 The Avoidance Problem

The rules of our type system (particularly rules 8, 9, and 11 from Figure 4) are careful to ensure that substituted modules are always determinate, at the expense of requiring that functor and second-projection arguments are determinate. This is necessary because the result of substituting an indeterminate module into a well-formed signature can be ill-formed. Thus, to apply a functor to an indeterminate argument, one must let-bind the argument and apply the functor to the resulting (determinate) variable.

A similar restriction is imposed by Shao [30], but Harper and Lillibridge [9] propose an alternative that softens the restriction. Harper and Lillibridge’s proposal (expressed in our terms) is to include a non-dependent typing rule without a determinacy restriction:

$$\frac{\Gamma \vdash_{\kappa} M_1 : \sigma_1 \rightarrow \sigma_2 \quad \Gamma \vdash_{\kappa} M_2 : \sigma_1}{\Gamma \vdash_{\kappa} M_1 M_2 : \sigma_2}$$

When M_2 is determinate, this rule carries the same force as our dependent rule, by exploiting singleton signatures and the contravariance of functor signatures:

$$\begin{aligned} \Pi s : \sigma_1 . \sigma_2 &\leq \Pi s : \mathfrak{S}_{\sigma_1}(M_2) . \sigma_2 \\ &\equiv \Pi s : \mathfrak{S}_{\sigma_1}(M_2) . \sigma_2[M_2/s] \\ &= \mathfrak{S}_{\sigma_1}(M_2) \rightarrow \sigma_2[M_2/s] \end{aligned}$$

When M_2 is indeterminate, this rule is more expressive than our typing rule, because the application can still occur. However, to exploit this rule, the type checker must find a non-dependent supersignature that is suitable for application to M_2 .

The *avoidance problem* [7, 16] is that there is no “best” way to do so. For example, consider the signature:

$$\sigma = (\llbracket T \rrbracket \rightarrow \mathfrak{S}(s)) \times \mathfrak{S}(s)$$

To obtain a supersignature of σ avoiding the variable s , we must forget that the first component is a constant function, and therefore we can only say that the second component is equal to the first component’s result on some particular argument. Thus, for any type τ , we may promote σ to the supersignature:

$$\Sigma F : (\llbracket T \rrbracket \rightarrow \llbracket T \rrbracket) . \mathfrak{S}(F[\tau])$$

This gives us an infinite array of choices. Any of these choices is superior to the obvious $(\llbracket T \rrbracket \rightarrow \llbracket T \rrbracket) \times \llbracket T \rrbracket$, but none of them is comparable to any other, since F is abstract. Thus, there is no minimal supersignature of σ avoiding s . The absence of minimal signatures is a problem, because it

means that there is no obvious way to perform type checking.

In our type system, we circumvent the avoidance problem by requiring that the arguments of functor application and second-projection be determinate (thereby eliminating any need to find non-dependent supersignatures), and provide a let construct so that such operations can still be applied to indeterminate modules. We have shown that, as a result, our type theory does enjoy principal signatures.

To achieve this, however, our let construct must be labelled with its result signature (not mentioning the variable being bound), for otherwise the avoidance problem re-arises. This essentially requires that every functor application or projection involving an indeterminate argument be labelled with its result signature as well, leading to potentially unacceptable syntactic overhead in practice. Fortunately, programs can be systematically rewritten to avoid this problem, as we describe next.

4.1 Elaboration and Existential Signatures

Consider the unannotated let expression $\text{let } s = M_1 \text{ in } M_2$, where $M_1 : \sigma_1$ and $M_2 : \sigma_2(s)$. If M_1 is determinate, then the let expression can be given the minimal signature $\sigma_2(M_1)$. Otherwise, we are left with the variable s leaving scope, but no minimal supersignature of $\sigma_2(s)$ not mentioning s . However, if we rewrite the let expression as the pair $(s = M_1, M_2)$, then we may give it the signature $\Sigma s : \sigma_1 . \sigma_2(s)$ and no avoidance problem arises. Similarly, the functor application $F(M)$ with $F : \Pi s : \sigma_1 . \sigma_2$ and indeterminate $M : \sigma_1$ can be rewritten as $\langle s = M, F(s) \rangle$ and given signature $\Sigma s : \sigma_1 . \sigma_2$.

Following Harper and Stone [11], we propose the use of an elaboration algorithm to systematize these rewritings. This elaborator takes code written in an external language that supports unannotated lets, as well as indeterminate functor application and second-projection, and produces code written in our type system. Since the elaborator rewrites modules in a manner that changes their signatures, it also must take responsibility for converting those modules back to their expected signature wherever required. This means that the elaborator must track which pairs are “real” and which have been invented by the elaborator to circumvent the avoidance problem.

The elaborator does so using the types. When the elaborator invents a pair to circumvent the avoidance problem, it gives its signature using an existential \exists rather than Σ . In the internal language, $\exists s : \sigma_1 . \sigma_2$ means the same thing as $\Sigma s : \sigma_1 . \sigma_2$, but the elaborator treats the two signatures differently: When the elaborator expects (say) a functor and encounters a $\Sigma s : \sigma_1 . \sigma_2$, it generates a type error. However, when it encounters an $\exists s : \sigma_1 . \sigma_2$, it extracts the σ_2 component (the elaborator’s invariants ensure that it always can do so), looking for the expected functor. Space considerations preclude further details of the elaboration algorithm, which appear in the companion technical report [6].

In a sense, the elaborator solves the avoidance problem by introducing existential signatures to serve in place of the non-existent minimal supersignatures not mentioning a variable. In light of this, a natural question is whether the need for an elaborator could be eliminated by making existential signatures primitive to the type system.

One natural way to govern primitive existentials is with

the introduction and elimination rules:

$$\frac{\Gamma \vdash_p M : \sigma_1 \quad \Gamma \vdash \sigma \leq \sigma_2[M/s] \quad \Gamma, s:\sigma_1 \vdash \sigma_2 \text{ sig}}{\Gamma \vdash \sigma \leq \exists s:\sigma_1.\sigma_2}$$

and

$$\frac{\Gamma, s:\sigma_1 \vdash \sigma_2 \leq \sigma \quad \Gamma \vdash \sigma_1 \text{ sig} \quad \Gamma \vdash \sigma \text{ sig}}{\Gamma \vdash \exists s:\sigma_1.\sigma_2 \leq \sigma}$$

With these rules, the avoidance problem could be solved: The least supersignature of $\sigma_2(s)$ not mentioning $s:\sigma_1$ would be $\exists s:\sigma_1.\sigma_2(s)$.

Unfortunately, these rules (particularly the first) make type checking undecidable. For example, each of the queries

$$\Pi s:\sigma. \llbracket \tau \rrbracket \stackrel{?}{\leq} \exists s':\sigma. \Pi s:\mathfrak{F}_\sigma(s'). \llbracket \tau' \rrbracket$$

and

$$(\lambda s:\sigma. \llbracket \tau \rrbracket) \stackrel{?}{\cong} (\lambda s:\sigma. \llbracket \tau' \rrbracket) : \exists s':\sigma. \Pi s:\mathfrak{F}_\sigma(s'). \llbracket T \rrbracket$$

holds if and only if there exists determinate $M : \sigma$ such that the types $\tau[M/s]$ and $\tau'[M/s]$ are equal. Thus, deciding subsignature or equivalence queries in the presence of existentials would be as hard as higher-order unification, which is known to be undecidable [8].

4.2 Syntactic Principal Signatures

It has been argued for reasons related to separate compilation that principal signatures should be expressible in the syntax available to the programmer. This provides the strongest support for separate compilation, because a programmer can break a program at any point and write an interface that expresses all the information the compiler could have determined at that point. Such strong support does not appear to be vital in practice, since systems such as Objective Caml and Standard ML of New Jersey’s higher-order modules have been used successfully for some time without principal signatures at all, but it is nevertheless a desirable property.

Our type system (*i.e.*, the internal language) does provide syntactic principal signatures, since principal signatures exist, and all the syntax is available to the programmer. However, the elaborator’s *external* language does not provide syntax for the existential signatures that can appear in elaborator signatures, which should be thought of as the principal signatures of external modules. Thus, we can say that our basic type system provides syntactic principal signatures, but our external language does not.

In an external language where the programmer is permitted to write existential signatures, elaborating code such as:

$$(\lambda s':(\exists s:\sigma_1.\sigma_2) \dots) M$$

requires the elaborator to decide whether M can be coerced to belong to $\exists s:\sigma_1.\sigma_2$, which in turn requires the elaborator to produce a $M' : \sigma_1$ such that $M : \sigma_2[M'/s]$. Determining whether any such M' exists requires the elaborator to solve an undecidable higher-order unification problem: if $\sigma_2 = \mathfrak{F}(\llbracket \tau \rrbracket) \rightarrow \mathfrak{F}(\llbracket \tau' \rrbracket)$ and $M = \lambda t:\llbracket T \rrbracket.t$, then $M : \sigma_2[M'/s]$ if and only if $\tau[M'/s]$ and $\tau'[M'/s]$ are equal.

Thus, to allow programmer-specified existential signatures in the greatest possible generality would make elaboration undecidable. Partial measures may be possible, but we will not discuss any here.

5 Packaging Modules as First-Class Values

It is desirable for modules to be usable as first-class values. This is useful to make it possible to choose at run time the most efficient implementation of a signature for a particular data set (for example, sparse or dense representations of arrays). Unfortunately, fully general first-class modules present difficulties for static typing [16].

One practical approach to modules as first-class values was suggested by Mitchell, *et al.* [23], who propose that second-class modules automatically be wrapped as existential packages [24] to obtain first-class values. A similar approach to modules as first-class values is described by Russo and implemented in Moscow ML [28].

This existential-packaging approach to modules as first-class values is built into our language. We write the type of a packaged module as $\langle \sigma \rangle$ and the packaging construct as `pack M as $\langle \sigma \rangle$` . Elimination of packaged modules (as for existentials) is performed using a closed-scope unpacking construct. These may be defined as follows:

$$\begin{aligned} \langle \sigma \rangle &\stackrel{\text{def}}{=} \forall \alpha. (\sigma \rightarrow \alpha) \rightarrow \alpha \\ \text{pack } M \text{ as } \langle \sigma \rangle &\stackrel{\text{def}}{=} \Lambda \alpha. \lambda f: (\sigma \rightarrow \alpha). f M \\ \text{unpack } e \text{ as } s:\sigma \text{ in } (e' : \tau) &\stackrel{\text{def}}{=} e \tau (\Lambda s:\sigma. e') \end{aligned}$$

(Compare the definition of $\langle \sigma \rangle$ with the standard encoding of the existential type $\exists \beta. \tau$ as $\forall \alpha. (\forall \beta. \tau \rightarrow \alpha) \rightarrow \alpha$.)

The main limitation of existentially-packaged modules is the closed-scope elimination construct. It has been observed repeatedly in the literature⁵ that this construct is too restrictive to be useful. For one, in “`unpack e as $s:\sigma$ in $(e' : \tau)$ ”, the result type τ may not mention s . As a consequence, functions over packaged modules may not be dependent; that is, the result type may not mention the argument. This deficiency is mitigated in our language by the ability to write functions over unpackaged, second-class modules, which can be given the dependent type $\Pi s:\sigma. \tau(s)$ instead of $\langle \sigma \rangle \rightarrow \tau$.`

Another problem with the closed-scope elimination construct is that a term of package type cannot be unpacked into a *stand-alone* second-class module; it can only be unpacked inside an enclosing term. As each unpacking of a packaged module creates an abstract type in a separate scope, packages must be unpacked at a very early stage to ensure coherence among their clients, leading to “scope inversions” that are awkward to manage in practice.

What we desire, therefore, is a new module construct of the form “`unpack e as σ` ”, which coerces a first-class package e of type $\langle \sigma \rangle$ back into a second-class module of signature σ . The following example illustrates how adding such a construct carelessly can lead to unsoundness:

```
module F = λs:⟨σ⟩.(unpack (Val s) as σ)
module X1 = F [pack M1 as ⟨σ⟩]
module X2 = F [pack M2 as ⟨σ⟩]
```

Note that the argument of the functor F is an atomic term module, so all arguments to F are statically equivalent. If F is given an applicative signature, then X_1 and X_2 will be deemed equivalent, even if the original modules M_1 and M_2 are not! Thus, F must be deemed generative, which in turn requires that the `unpack` construct induce a *dynamic* effect.

⁵originally by MacQueen [18] and later by Cardelli and Leroy [3] and Lillibridge [16], among others

$$\begin{array}{lcl}
\text{types} & \tau ::= \dots & | \langle \sigma \rangle \\
\text{terms} & e ::= \dots & | \text{pack } M \text{ as } \langle \sigma \rangle \\
\text{modules} & M ::= \dots & | \text{unpack } e \text{ as } \sigma
\end{array}$$

$$\frac{\Gamma \vdash \sigma_1 \equiv \sigma_2}{\Gamma \vdash \langle \sigma_1 \rangle \equiv \langle \sigma_2 \rangle} \quad \frac{\Gamma \vdash_{\kappa} M : \sigma}{\Gamma \vdash \text{pack } M \text{ as } \langle \sigma \rangle : \langle \sigma \rangle}$$

$$\frac{\Gamma \vdash e : \langle \sigma \rangle}{\Gamma \vdash_s \text{unpack } e \text{ as } \sigma : \sigma}$$

Figure 8: Packaged Module Extension

Packaged modules that admit this improved unpacking construct are not definable in our core language, but they constitute a simple, orthogonal extension to the type system that does not complicate type checking. The syntax and typing rules for this extension are given in Figure 8. Note that the closed-scope unpacking construct is definable as

$$\text{let } s = (\text{unpack } e \text{ as } \sigma) \text{ in } (e' : \tau)$$

Intuitively, unpacking is generative because the module being unpacked can be an arbitrary term, whose type components may depend on run-time conditions. In the core system we presented in Section 2, the generativity induced by strong sealing was merely a *pro forma* effect—the language, supporting only second-class modules, provided no way for the type components of a module to be actually generated at run time. The type system, however, treats dynamic effects as if they are all truly dynamic, and thus it scales easily to handle the real run-time type generation enabled by the extension in Figure 8.

6 Related Work

Harper, Mitchell and Moggi [10] pioneered the theory of *phase separation*, which is fundamental to achieving maximal type propagation in higher-order module systems. Their non-standard equational rules, which identify higher-order modules with primitive “phase-split” ones, are similar in spirit to, though different in detail from, our notion of static module equivalence. One may view their system as a subsystem of ours in which there is no sealing mechanism (and consequently all modules are determinate).

MacQueen and Tofte [19] proposed a higher-order module extension to the original Definition of Standard ML [21], which was implemented in the Standard ML of New Jersey compiler. Their semantics involves a two-phase elaboration process, in which higher-order functors are re-elaborated at each application to take advantage of additional information about their arguments. This advantage is balanced by the disadvantage of inhibiting type propagation in the presence of separate compilation since functors that are compiled separately from their applications cannot be re-elaborated. A more thorough comparison is difficult because MacQueen and Tofte employ a stamp-based semantics, which is difficult to transfer to a type-theoretic setting.

Focusing on controlled abstraction, but largely neglecting higher-order modules, Harper and Lillibridge [9] and Leroy [12, 14] introduced the closely related concepts of *translucent sums* and *manifest types*. These mechanisms served as the basis of the module system in the revised Definition of Standard ML 1997 [22], and Harper and Stone [11]

have formalized the elaboration of Standard ML 1997 programs into a translucent sums calculus. To deal with the avoidance problem, Harper and Stone rely on elaborator mechanisms similar to ours. The Harper and Stone language can be viewed as a subsystem of ours in which all functors are generative and only strong sealing is supported.

Leroy introduced the notion of an *applicative functor* [13], which enables one to give fully transparent signatures for many higher-order functors. Leroy’s formalism defined determinacy by a syntactic restriction that functor applications appearing in type paths must be in named form. On one hand, this restriction provides a weak form of structure sharing in the sense that the abstract type $F(X).t$ can only be the result of applying F to the module *named* X . On the other hand, the restriction prevents the system from capturing the full equational theory of higher-order functors, since not all equations can be expressed in named form [4].

Together, manifest types and applicative functors form the basis of the module system of Objective Caml [26]. The manifest type formalism, like the translucent sum formalism, does not address the avoidance problem, and consequently it lacks principal signatures. Aside from this fact (and disregarding Objective Caml’s signature fields in structures, which make typechecking undecidable [16]), the module language of Objective Caml can be viewed as essentially a subsystem of Russo’s system (below).

More recently, Russo [27] has proposed a type system for applicative functors that generalizes Leroy’s in much the same way we do, by abandoning the named form restriction. We adopt his use of existential signatures to address the avoidance problem, although Russo simultaneously uses existentials to model generativity, which we do not. Russo has implemented his system as an experimental extension to the Moscow ML compiler [25], which already supports SML-style generative functors. One can view Moscow ML’s module system as a subsystem of ours in which weak sealing is the only available form of abstraction and in which all modules are considered dynamically pure. As we observed in Section 3, a system with only dynamically pure modules can provide a faithful account of generativity only in the absence of functors. Thus, the introduction of applicative functors into Moscow ML has the effect of making its generative functors weaker, since one can defeat generativity by eta-expanding a generative functor into an applicative one. Russo has also implemented an extension to Moscow ML for packaging modules as first-class values [28]. His extension is very similar to the existential-packaging approach discussed in the beginning of Section 5, and suffers likewise from the limitations of the closed-scope unpacking construct.

Shao [30] proposes another type system for modules supporting both applicative and generative functors. Shao’s system, in contrast to Russo’s, may be viewed as a subsystem of ours based exclusively on strong sealing instead of weak sealing. As we observed in Section 3, a consequence of this is that the bodies of applicative functors may not contain any opaque substructures, such as (opaque) datatypes. Shao’s system, like ours (recall Section 4), circumvents the avoidance problem by restricting functor application and projection to determinate arguments (which must be in named form in his system), and by eliminating implicit subsumption (in essence, this requires let expressions to be annotated, as in our system). We conjecture that our elaboration techniques could be applied to Shao’s system to lift these restrictions in his system as well, at the expense of syntactic principal signatures.

7 Conclusion

Type systems for first-order module systems are reasonably well understood. In contrast, previous work on type-theoretic, higher-order modules has left that field in a fragmented state, with various competing designs and no clear statement of the trade-offs (if any) between those designs. This state of the field has made it difficult to choose one design over another, and has left the erroneous impression of trade-offs that do not actually exist. For example, no previous design supports both (undefeatable) generativity and applicative functors with opaque subcomponents.

Our language seeks to unify the field by providing a practical type system for higher-order modules that simultaneously supports the key functionality of preceding module systems. In the process we dispel some misconceptions, such as a trade-off between fully expressive generative and applicative functors, thereby eliminating some dilemmas facing language designers.

Nevertheless, there are several important issues in modular programming that go beyond the scope of our type theory, including:

- *Structure Sharing.* The original version of Standard ML [21] included a notion of module equivalence that was sensitive to the dynamic, as well as static, parts of the module. Since this would violate our fundamental commitment to phase separation, it appears difficult to give an intrinsic type-theoretic account of this notion. It is in fact possible to simulate structure sharing by having the elaborator add an abstract type to each structure to serve as the “compile-time name” of that structure. But this would be merely an elaboration convention, not a true type-theoretic account of structure sharing.
- *Recursive Modules.* An important direction for future research is to integrate recursive modules [5] into the present framework. The chief difficulty is to achieve decidable type checking in the presence of general recursively dependent signatures, or to isolate a practical sub-language that avoids these problems. Russo [29] has formalized an extension to Moscow ML along these lines, but it suffers from weak support for separate compilation.
- *Extensibility.* A great deal of research has been devoted to incorporating forms of extensible programming, common in object-oriented settings, into functional languages. Many of the resulting languages include some kind of class mechanism [26], whose type-theoretic interpretation is often complex. Others make use of multimethods, for which properties such as redundancy and exhaustiveness are difficult to track in the presence of modules [20]. Understanding how these different forms of extensibility relate to one another, and how they interact with higher-order modules, is a substantial problem for future work.

References

- [1] David R. Aspinall. *Type Systems for Modular Programs and Specifications*. PhD thesis, Edinburgh University, Edinburgh, Scotland, December 1997.
- [2] Luca Cardelli, Jim Donahue, Mick Jordan, Bill Kalso, and Greg Nelson. The Modula-3 type system. In *Sixteenth ACM Symposium on Principles of Programming Languages*, pages 202–212, Austin, TX, January 1989.
- [3] Luca Cardelli and Xavier Leroy. Abstract types and the dot notation. In M. Broy and C. B. Jones, editors, *Proceedings IFIP TC2 working conference on programming concepts and methods*, pages 479–504. North-Holland, 1990. Also available as research report 56, DEC Systems Research Center.
- [4] Karl Crary. Sound and complete elimination of singleton kinds. In *Third Workshop on Types in Compilation*, volume 2071 of *Lecture Notes in Computer Science*, pages 1–25. Springer-Verlag, September 2000. Extended version published as CMU technical report CMU-CS-00-104.
- [5] Karl Crary, Robert Harper, and Sidd Puri. What is a recursive module? In *SIGPLAN '99 Conference on Programming Language Design and Implementation (PLDI)*, pages 50–63, Atlanta, GA, 1999. ACM SIGPLAN.
- [6] Derek Dreyer, Karl Crary, and Robert Harper. A type system for higher-order modules. Technical Report CMU-CS-02-122, School of Computer Science, Carnegie Mellon University, March 2002.
- [7] Giorgio Ghelli and Benjamin Pierce. Bounded existentials and minimal typing. *Theoretical Computer Science*, 193:75–96, 1998.
- [8] Warren D. Goldfarb. The undecidability of the second-order unification problem. *Theoretical Computer Science*, 13:225–230, 1981.
- [9] Robert Harper and Mark Lillibridge. A type-theoretic approach to higher-order modules with sharing. In *Twenty-First ACM Symposium on Principles of Programming Languages*, pages 123–137, Portland, OR, January 1994.
- [10] Robert Harper, John C. Mitchell, and Eugenio Moggi. Higher-order modules and the phase distinction. In *Seventeenth ACM Symposium on Principles of Programming Languages*, San Francisco, CA, January 1990.
- [11] Robert Harper and Chris Stone. A type-theoretic interpretation of Standard ML. In Gordon Plotkin, Colin Stirling, and Mads Tofte, editors, *Proof, Language, and Interaction: Essays in Honor of Robin Milner*. MIT Press, 2000.
- [12] Xavier Leroy. Manifest types, modules, and separate compilation. In *Proceedings of the Twenty-first Annual ACM Symposium on Principles of Programming Languages, Portland*. ACM, January 1994.

- [13] Xavier Leroy. Applicative functors and fully transparent higher-order modules. In *Conference Record of POPL '95: ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, pages 142–153, San Francisco, CA, January 1995.
- [14] Xavier Leroy. A syntactic theory of type generativity and sharing. *Journal of Functional Programming*, 6(5):667–698, 1996.
- [15] Xavier Leroy. A modular module system. *Journal of Functional Programming*, 10(3):269–303, 2000.
- [16] Mark Lillibridge. *Translucent Sums: A Foundation for Higher-Order Module Systems*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, December 1996.
- [17] Barbara Liskov and John Guttag. *Abstraction and Specification in Program Development*. MIT Press, 1986.
- [18] David MacQueen. Using dependent types to express modular structure. In *Thirteenth ACM Symposium on Principles of Programming Languages*, 1986.
- [19] David B. MacQueen and Mads Tofte. A semantics for higher-order functors. In Donald T. Sannella, editor, *Programming Languages and Systems — ESOP '94*, volume 788 of *Lecture Notes in Computer Science*, pages 409–423. Springer-Verlag, 1994.
- [20] Todd Millstein and Craig Chambers. Modular type-checking for hierarchically extensible datatypes and functions. In *9th Workshop on Foundations of Object-Oriented Languages*, pages 1–12, Portland, Oregon, January 2002.
- [21] Robin Milner, Mads Tofte, and Robert Harper. *The Definition of Standard ML*. MIT Press, 1990.
- [22] Robin Milner, Mads Tofte, Robert Harper, and David MacQueen. *The Definition of Standard ML (Revised)*. MIT Press, 1997.
- [23] John Mitchell, Sigurd Meldal, and Neel Madhav. An extension of Standard ML modules with subtyping and inheritance. In *Eighteenth ACM Symposium on Principles of Programming Languages*, January 1991.
- [24] John C. Mitchell and Gordon D. Plotkin. Abstract types have existential type. *ACM Transactions on Programming Languages and Systems*, 10(3):470–502, July 1988.
- [25] Moscow ML. <http://www.dina.kvl.dk/~sestoft/mosml.html>.
- [26] Objective Caml. <http://www.ocaml.org>.
- [27] Claudio Russo. *Types for Modules*. PhD thesis, Edinburgh University, Edinburgh, Scotland, 1998. LFCS Thesis ECS-LFCS-98-389.
- [28] Claudio V. Russo. First-class structures for Standard ML. *Nordic Journal of Computing*, 7(4):348–374, 2000.
- [29] Claudio V. Russo. Recursive structures for Standard ML. In *International Conference on Functional Programming*, pages 50–61, Florence, Italy, September 2001.
- [30] Zhong Shao. Transparent modules with fully syntactic signatures. In *International Conference on Functional Programming*, pages 220–232, Paris, France, September 1999.
- [31] Christopher A. Stone. *Singleton Kinds and Singleton Types*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, August 2000.
- [32] Christopher A. Stone and Robert Harper. Deciding type equivalence in a language with singleton kinds. In *Twenty-Seventh ACM Symposium on Principles of Programming Languages*, pages 214–227, Boston, January 2000.
- [33] Niklaus Wirth. *Programming in Modula-2*. Texts and Monographs in Computer Science. Springer-Verlag, 1983.

Notes on Appendices

Appendix A gives the type system for our module calculus. Appendix B gives the module typechecking and principal signature synthesis judgements that form the core of our typechecking algorithm. We omit the judgments for term typechecking ($\Gamma \vdash e \Leftarrow \tau$) and unique type synthesis ($\Gamma \vdash e \Rightarrow \tau$) for space reasons; they are fully detailed in the companion technical report [6].

A Type System

Well-formed contexts: $\Gamma \vdash \text{ok}$

$$\frac{}{\epsilon \vdash \text{ok}} \quad \frac{\Gamma \vdash \sigma \text{ sig} \quad s \notin \text{dom}(\Gamma)}{\Gamma, s:\sigma \vdash \text{ok}}$$

Well-formed types: $\Gamma \vdash \tau \text{ type}$

$$\frac{\Gamma \vdash_{\text{P}} M : [T]}{\Gamma \vdash \text{Typ } M \text{ type}} \quad \frac{\Gamma, s:\sigma \vdash \tau \text{ type}}{\Gamma \vdash \Pi s:\sigma.\tau \text{ type}} \quad \frac{\Gamma \vdash \tau' \text{ type} \quad \Gamma \vdash \tau'' \text{ type}}{\Gamma \vdash \tau' \times \tau'' \text{ type}} \quad \frac{\Gamma \vdash \sigma \text{ sig}}{\Gamma \vdash \langle \sigma \rangle \text{ type}}$$

Type equivalence: $\Gamma \vdash \tau_1 \equiv \tau_2$

$$\frac{\Gamma \vdash [\tau_1] \cong [\tau_2] : [T]}{\Gamma \vdash \tau_1 \equiv \tau_2} \quad \frac{\Gamma \vdash \sigma_1 \equiv \sigma_2 \quad \Gamma, s:\sigma_1 \vdash \tau_1 \equiv \tau_2}{\Gamma \vdash \Pi s:\sigma_1.\tau_1 \equiv \Pi s:\sigma_2.\tau_2} \quad \frac{\Gamma \vdash \tau'_1 \equiv \tau'_2 \quad \Gamma \vdash \tau''_1 \equiv \tau''_2}{\Gamma \vdash \tau'_1 \times \tau''_1 \equiv \tau'_2 \times \tau''_2} \quad \frac{\Gamma \vdash \sigma_1 \equiv \sigma_2}{\Gamma \vdash \langle \sigma_1 \rangle \equiv \langle \sigma_2 \rangle}$$

Well-formed terms: $\Gamma \vdash e : \tau$

$$\frac{\Gamma \vdash e : \tau' \quad \Gamma \vdash \tau' \equiv \tau}{\Gamma \vdash e : \tau} \quad \frac{\Gamma \vdash_{\kappa} M : [\tau]}{\Gamma \vdash \text{Val } M : \tau} \quad \frac{\Gamma \vdash_{\kappa} M : \sigma \quad \Gamma, s:\sigma \vdash e : \tau \quad \Gamma \vdash \tau \text{ type}}{\Gamma \vdash \text{let } s = M \text{ in } (e : \tau) : \tau} \quad \frac{\Gamma \vdash_{\kappa} M : \sigma}{\Gamma \vdash \text{pack } M \text{ as } \langle \sigma \rangle : \langle \sigma \rangle}$$

$$\frac{\Gamma, f:[\Pi s:\sigma.\tau], s:\sigma \vdash e : \tau}{\Gamma \vdash \text{fix } f(s:\sigma):\tau.e : \Pi s:\sigma.\tau} \quad \frac{\Gamma \vdash e : \Pi s:\sigma.\tau \quad \Gamma \vdash_{\text{P}} M : \sigma}{\Gamma \vdash e M : \tau[M/s]} \quad \frac{\Gamma \vdash e' : \tau' \quad \Gamma \vdash e'' : \tau''}{\Gamma \vdash \langle e', e'' \rangle : \tau' \times \tau''} \quad \frac{\Gamma \vdash e : \tau' \times \tau''}{\Gamma \vdash \pi_1 e : \tau'} \quad \frac{\Gamma \vdash e : \tau' \times \tau''}{\Gamma \vdash \pi_2 e : \tau''}$$

Well-formed signatures: $\Gamma \vdash \sigma \text{ sig}$

$$\frac{\Gamma \vdash \text{ok}}{\Gamma \vdash 1 \text{ sig}} \quad \frac{\Gamma \vdash \text{ok}}{\Gamma \vdash [T] \text{ sig}} \quad \frac{\Gamma \vdash \tau \text{ type}}{\Gamma \vdash [\tau] \text{ sig}} \quad \frac{\Gamma \vdash_{\text{P}} M : [T]}{\Gamma \vdash \mathfrak{S}(M) \text{ sig}} \quad \frac{\Gamma, s:\sigma' \vdash \sigma'' \text{ sig}}{\Gamma \vdash \Pi^{\delta} s:\sigma'.\sigma'' \text{ sig}} \quad \frac{\Gamma, s:\sigma' \vdash \sigma'' \text{ sig}}{\Gamma \vdash \Sigma s:\sigma'.\sigma'' \text{ sig}}$$

Signature equivalence: $\Gamma \vdash \sigma_1 \equiv \sigma_2$

$$\frac{\Gamma \vdash \text{ok}}{\Gamma \vdash 1 \equiv 1} \quad \frac{\Gamma \vdash \text{ok}}{\Gamma \vdash [T] \equiv [T]} \quad \frac{\Gamma \vdash \tau_1 \equiv \tau_2}{\Gamma \vdash [\tau_1] \equiv [\tau_2]} \quad \frac{\Gamma \vdash M_1 \cong M_2 : [T]}{\Gamma \vdash \mathfrak{S}(M_1) \equiv \mathfrak{S}(M_2)}$$

$$\frac{\Gamma \vdash \sigma'_2 \equiv \sigma'_1 \quad \Gamma, s:\sigma'_2 \vdash \sigma''_1 \equiv \sigma''_2}{\Gamma \vdash \Pi^{\delta} s:\sigma'_1.\sigma''_1 \equiv \Pi^{\delta} s:\sigma'_2.\sigma''_2} \quad \frac{\Gamma \vdash \sigma'_1 \equiv \sigma'_2 \quad \Gamma, s:\sigma'_1 \vdash \sigma''_1 \equiv \sigma''_2}{\Gamma \vdash \Sigma s:\sigma'_1.\sigma''_1 \equiv \Sigma s:\sigma'_2.\sigma''_2}$$

Signature subtyping: $\Gamma \vdash \sigma_1 \leq \sigma_2$

$$\frac{\Gamma \vdash \text{ok}}{\Gamma \vdash 1 \leq 1} \quad \frac{\Gamma \vdash \text{ok}}{\Gamma \vdash [T] \leq [T]} \quad \frac{\Gamma \vdash \tau_1 \equiv \tau_2}{\Gamma \vdash [\tau_1] \leq [\tau_2]} \quad \frac{\Gamma \vdash_{\text{P}} M : [T]}{\Gamma \vdash \mathfrak{S}(M) \leq [T]} \quad \frac{\Gamma \vdash M_1 \cong M_2 : [T]}{\Gamma \vdash \mathfrak{S}(M_1) \leq \mathfrak{S}(M_2)}$$

$$\frac{\Gamma \vdash \sigma'_2 \leq \sigma'_1 \quad \Gamma, s:\sigma'_2 \vdash \sigma''_1 \leq \sigma''_2 \quad \Gamma, s:\sigma'_1 \vdash \sigma''_1 \text{ sig} \quad \delta_1 \sqsubseteq \delta_2}{\Gamma \vdash \Pi^{\delta_1} s:\sigma'_1.\sigma''_1 \leq \Pi^{\delta_2} s:\sigma'_2.\sigma''_2} \quad \frac{\Gamma \vdash \sigma'_1 \leq \sigma'_2 \quad \Gamma, s:\sigma'_1 \vdash \sigma''_1 \leq \sigma''_2 \quad \Gamma, s:\sigma'_2 \vdash \sigma''_2 \text{ sig}}{\Gamma \vdash \Sigma s:\sigma'_1.\sigma''_1 \leq \Sigma s:\sigma'_2.\sigma''_2}$$

Well-formed modules: $\Gamma \vdash_{\kappa} M : \sigma$

$$\frac{\Gamma \vdash \text{ok}}{\Gamma \vdash_{\text{P}} s : \Gamma(s)} \quad \frac{\Gamma \vdash \text{ok}}{\Gamma \vdash_{\text{P}} \langle \rangle : 1} \quad \frac{\Gamma \vdash \tau \text{ type}}{\Gamma \vdash_{\text{P}} [\tau] : [T]} \quad \frac{\Gamma \vdash e : \tau \quad \Gamma \vdash \tau \text{ type}}{\Gamma \vdash_{\text{P}} [e : \tau] : [\tau]} \quad \frac{\Gamma, s:\sigma' \vdash_{\kappa} M : \sigma'' \quad \kappa \sqsubseteq \text{D}}{\Gamma \vdash_{\kappa} \lambda s:\sigma'.M : \Pi s:\sigma'.\sigma''}$$

$$\frac{\Gamma, s:\sigma' \vdash_{\kappa} M : \sigma'' \quad \Gamma, s:\sigma' \vdash \sigma'' \text{ sig}}{\Gamma \vdash_{\kappa \sqcap \text{D}} \lambda s:\sigma'.M : \Pi^{\text{gen}} s:\sigma'.\sigma''} \quad \frac{\Gamma \vdash_{\kappa} F : \Pi s:\sigma'.\sigma'' \quad \Gamma \vdash_{\text{P}} M : \sigma'}{\Gamma \vdash_{\kappa} FM : \sigma''[M/s]} \quad \frac{\Gamma \vdash_{\kappa} F : \Pi^{\text{gen}} s:\sigma'.\sigma'' \quad \Gamma \vdash_{\text{P}} M : \sigma'}{\Gamma \vdash_{\kappa \sqcup \text{S}} FM : \sigma''[M/s]}$$

$$\frac{\Gamma \vdash_{\kappa} M' : \sigma' \quad \Gamma, s:\sigma' \vdash_{\kappa} M'' : \sigma''}{\Gamma \vdash_{\kappa} \langle s = M', M'' \rangle : \Sigma s:\sigma'.\sigma''} \quad \frac{\Gamma \vdash_{\kappa} M : \Sigma s:\sigma'.\sigma''}{\Gamma \vdash_{\kappa} \pi_1 M : \sigma'} \quad \frac{\Gamma \vdash_{\text{P}} M : \Sigma s:\sigma'.\sigma''}{\Gamma \vdash_{\text{P}} \pi_2 M : \sigma''[\pi_1 M/s]} \quad \frac{\Gamma \vdash_{\text{P}} M : [T]}{\Gamma \vdash_{\text{P}} M : \mathfrak{S}(M)}$$

$$\frac{\Gamma \vdash_{\kappa} M : \sigma}{\Gamma \vdash_{\kappa \sqcup \text{D}} (M :: \sigma) : \sigma} \quad \frac{\Gamma \vdash_{\kappa} M : \sigma}{\Gamma \vdash_{\text{W}} (M :> \sigma) : \sigma} \quad \frac{\Gamma, s:\sigma' \vdash_{\text{P}} Ms : \sigma'' \quad \Gamma \vdash_{\text{P}} M : \Pi s:\sigma'.\rho}{\Gamma \vdash_{\text{P}} M : \Pi s:\sigma'.\sigma''} \quad \frac{\Gamma \vdash_{\text{P}} \pi_1 M : \sigma' \quad \Gamma \vdash_{\text{P}} \pi_2 M : \sigma''}{\Gamma \vdash_{\text{P}} M : \sigma' \times \sigma''}$$

$$\frac{\Gamma \vdash_{\kappa} M' : \sigma' \quad \Gamma, s:\sigma' \vdash_{\kappa} M'' : \sigma \quad \Gamma \vdash \sigma \text{ sig}}{\Gamma \vdash_{\kappa} \text{let } s = M' \text{ in } (M'' : \sigma) : \sigma} \quad \frac{\Gamma \vdash e : \langle \sigma \rangle}{\Gamma \vdash_{\text{S}} \text{unpack } e \text{ as } \sigma : \sigma} \quad \frac{\Gamma \vdash_{\kappa'} M : \sigma' \quad \Gamma \vdash \sigma' \leq \sigma \quad \kappa' \sqsubseteq \kappa}{\Gamma \vdash_{\kappa} M : \sigma}$$

Module equivalence: $\Gamma \vdash M_1 \cong M_2 : \sigma$

$$\begin{array}{c}
\frac{\Gamma \vdash_P M : \sigma}{\Gamma \vdash M \cong M : \sigma} \quad \frac{\Gamma \vdash M_2 \cong M_1 : \sigma}{\Gamma \vdash M_1 \cong M_2 : \sigma} \quad \frac{\Gamma \vdash M_1 \cong M_2 : \sigma \quad \Gamma \vdash M_2 \cong M_3 : \sigma}{\Gamma \vdash M_1 \cong M_3 : \sigma} \quad \frac{\Gamma \vdash \tau_1 \equiv \tau_2}{\Gamma \vdash [\tau_1] \cong [\tau_2] : [\mathcal{T}]} \quad \frac{\Gamma \vdash_P M : [\mathcal{T}]}{\Gamma \vdash [\text{Typ } M] \cong M : [\mathcal{T}]} \\
\\
\frac{\Gamma \vdash_P M_1 : 1 \quad \Gamma \vdash_P M_2 : 1}{\Gamma \vdash M_1 \cong M_2 : 1} \quad \frac{\Gamma \vdash_P M_1 : [\tau] \quad \Gamma \vdash_P M_2 : [\tau]}{\Gamma \vdash M_1 \cong M_2 : [\tau]} \quad \frac{\Gamma \vdash_P M_1 : \Pi^{\text{gen}} s : \sigma' . \sigma'' \quad \Gamma \vdash_P M_2 : \Pi^{\text{gen}} s : \sigma' . \sigma''}{\Gamma \vdash M_1 \cong M_2 : \Pi^{\text{gen}} s : \sigma' . \sigma''} \\
\\
\frac{\Gamma \vdash \sigma'_1 \equiv \sigma'_2 \quad \Gamma, s : \sigma'_1 \vdash M_1 \cong M_2 : \sigma''}{\Gamma \vdash \lambda s : \sigma'_1 . M_1 \cong \lambda s : \sigma'_2 . M_2 : \Pi s : \sigma'_1 . \sigma''} \quad \frac{\Gamma \vdash F_1 \equiv F_2 : \Pi s : \sigma' . \sigma'' \quad \Gamma \vdash M_1 \cong M_2 : \sigma'}{\Gamma \vdash F_1 M_1 \cong F_2 M_2 : \sigma''[M_1/s]} \\
\\
\frac{\Gamma \vdash M'_1 \cong M'_2 : \sigma' \quad \Gamma, s : \sigma' \vdash M''_1 \cong M''_2 : \sigma''}{\Gamma \vdash \langle s = M'_1, M''_1 \rangle \cong \langle s = M'_2, M''_2 \rangle : \Sigma s : \sigma' . \sigma''} \quad \frac{\Gamma \vdash M_1 \cong M_2 : \Sigma s : \sigma' . \sigma''}{\Gamma \vdash \pi_1 M_1 \cong \pi_1 M_2 : \sigma'} \quad \frac{\Gamma \vdash M_1 \cong M_2 : \Sigma s : \sigma' . \sigma''}{\Gamma \vdash \pi_2 M_1 \cong \pi_2 M_2 : \sigma''[\pi_1 M_1/s]} \\
\\
\frac{\Gamma, s : \sigma' \vdash M_1 s \cong M_2 s : \sigma'' \quad \Gamma \vdash_P M_1 : \Pi s : \sigma' . \rho_1 \quad \Gamma \vdash_P M_2 : \Pi s : \sigma' . \rho_2}{\Gamma \vdash M_1 \cong M_2 : \Pi s : \sigma' . \sigma''} \quad \frac{\Gamma \vdash \pi_1 M_1 \cong \pi_1 M_2 : \sigma' \quad \Gamma \vdash \pi_2 M_1 \cong \pi_2 M_2 : \sigma''}{\Gamma \vdash M_1 \cong M_2 : \sigma' \times \sigma''} \\
\\
\frac{\Gamma \vdash_P M' : \sigma' \quad \Gamma, s : \sigma' \vdash_P M'' : \sigma \quad \Gamma \vdash \sigma \text{ sig}}{\Gamma \vdash \text{let } s = M' \text{ in } (M'' : \sigma) \cong M''[M'/s] : \sigma} \quad \frac{\Gamma \vdash_P M_1 : \mathfrak{S}(M_2)}{\Gamma \vdash M_1 \cong M_2 : \mathfrak{S}(M_2)} \quad \frac{\Gamma \vdash M_1 \cong M_2 : \sigma' \quad \Gamma \vdash \sigma' \leq \sigma}{\Gamma \vdash M_1 \cong M_2 : \sigma}
\end{array}$$

B Typechecking Algorithm

Module typechecking: $\Gamma \vdash_{\kappa} M \Leftarrow \sigma$

$$\frac{\Gamma \vdash_{\kappa} M \Rightarrow \sigma' \quad \Gamma \vdash \sigma' \leq \sigma}{\Gamma \vdash_{\kappa} M \Leftarrow \sigma}$$

Principal signature synthesis: $\Gamma \vdash_{\kappa} M \Rightarrow \sigma$

$$\begin{array}{c}
\frac{\Gamma \vdash \text{ok}}{\Gamma \vdash_P s \Rightarrow \mathfrak{S}_{\Gamma(s)}(s)} \quad \frac{\Gamma \vdash \text{ok}}{\Gamma \vdash_P \langle \rangle \Rightarrow 1} \quad \frac{\Gamma \vdash \tau \text{ type}}{\Gamma \vdash_P [\tau] \Rightarrow \mathfrak{S}([\tau])} \quad \frac{\Gamma \vdash e \Leftarrow \tau}{\Gamma \vdash_P [e : \tau] \Rightarrow [\tau]} \quad \frac{\Gamma, s : \sigma' \vdash_{\kappa} M \Rightarrow \sigma'' \quad \kappa \sqsubseteq \mathcal{D}}{\Gamma \vdash_{\kappa} \lambda s : \sigma' . M \Rightarrow \Pi s : \sigma' . \sigma''} \\
\\
\frac{\Gamma, s : \sigma' \vdash_{\kappa} M \Rightarrow \sigma'' \quad \mathfrak{S} \sqsubseteq \kappa}{\Gamma \vdash_{\kappa \sqcap \mathcal{D}} \lambda s : \sigma' . M \Rightarrow \Pi^{\text{gen}} s : \sigma' . \sigma''} \quad \frac{\Gamma \vdash_{\kappa} F \Rightarrow \Pi s : \sigma' . \sigma'' \quad \Gamma \vdash_P M \Leftarrow \sigma'}{\Gamma \vdash_{\kappa} FM \Rightarrow \sigma''[M/s]} \quad \frac{\Gamma \vdash_{\kappa} F \Rightarrow \Pi^{\text{gen}} s : \sigma' . \sigma'' \quad \Gamma \vdash_P M \Leftarrow \sigma'}{\Gamma \vdash_{\kappa \sqcup \mathcal{S}} FM \Rightarrow \sigma''[M/s]} \\
\\
\frac{\Gamma \vdash_P M' \Rightarrow \sigma' \quad \Gamma, s : \sigma' \vdash_P M'' \Rightarrow \sigma''}{\Gamma \vdash_P \langle s = M', M'' \rangle \Rightarrow \sigma' \times \sigma''[M'/s]} \quad \frac{\Gamma \vdash_{\kappa'} M' \Rightarrow \sigma' \quad \Gamma, s : \sigma' \vdash_{\kappa''} M'' \Rightarrow \sigma'' \quad \kappa' \sqcup \kappa'' \neq \mathcal{P}}{\Gamma \vdash_{\kappa' \sqcup \kappa''} \langle s = M', M'' \rangle \Rightarrow \Sigma s : \sigma' . \sigma''} \\
\\
\frac{\Gamma \vdash_{\kappa} M \Rightarrow \Sigma s : \sigma' . \sigma''}{\Gamma \vdash_{\kappa} \pi_1 M \Rightarrow \sigma'} \quad \frac{\Gamma \vdash_P M \Rightarrow \sigma' \times \sigma''}{\Gamma \vdash_P \pi_2 M \Rightarrow \sigma''} \quad \frac{\Gamma \vdash_{\kappa} M \Leftarrow \sigma}{\Gamma \vdash_{\kappa \sqcup \mathcal{D}} M :: \sigma \Rightarrow \sigma} \quad \frac{\Gamma \vdash_{\kappa} M \Leftarrow \sigma}{\Gamma \vdash_{\mathcal{W}} M : > \sigma \Rightarrow \sigma} \quad \frac{\Gamma \vdash e \Leftarrow \langle \sigma \rangle}{\Gamma \vdash_{\mathcal{S}} \text{unpack } e \text{ as } \sigma \Rightarrow \sigma} \\
\\
\frac{\Gamma \vdash_P M' \Rightarrow \sigma' \quad \Gamma, s : \sigma' \vdash_P M'' \Leftarrow \sigma \quad \Gamma \vdash \sigma \text{ sig}}{\Gamma \vdash_P \text{let } s = M' \text{ in } (M'' : \sigma) \Rightarrow \mathfrak{S}_{\sigma}(\text{let } s = M' \text{ in } (M'' : \sigma))} \quad \frac{\Gamma \vdash_{\kappa'} M' \Rightarrow \sigma' \quad \Gamma, s : \sigma' \vdash_{\kappa''} M'' \Leftarrow \sigma \quad \Gamma \vdash \sigma \text{ sig} \quad \kappa' \sqcup \kappa'' \neq \mathcal{P}}{\Gamma \vdash_{\kappa' \sqcup \kappa''} \text{let } s = M' \text{ in } (M'' : \sigma) \Rightarrow \sigma}
\end{array}$$