

Phase Distinctions in Type Theory

Robert Harper
Carnegie Mellon University

(Joint work with Jon Sterling, Yue Niu, and Harrison Grodin)

December 2021

Acknowledgments

Thank you to David and Tim for the kind invitation!

Please see cited papers for references and discussion of related work.

This work was supported in part by **AFOSR** under grants MURI FA9550-15-1-0053 and FA9550-19-1-0216 (Tristan Nguyen, program manager) and in part by **NSF** under award number CCF-1901381. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the AFOSR or the NSF.

The Original Phase Distinction

Most PL's distinguish two **phases** of processing:

- **Compile-time**: parsing, type checking, compilation.
- **Run-time**: (compilation and) execution, including effects.

The distinction is **fundamental** to well-established development practices!

- Separate development.
- Interfaces to libraries.
- **Stability under change**.

Static vs Dynamic in PL's

The phase distinction is expressed by distinguishing **static** from **dynamic** levels.

Static part: **kinds** classify **constructors**

- Type classifies types.
- Closed under products, functions, singletons.

Dynamic part: **types** classify **code**.

- Products, sums, functions.
- Control and storage effects.

Dynamic part **depends on** static part, but static depends **only** on static.

Modules Are Mixed-Phase

Program modules **consolidate** static and dynamic parts.

```
signature QUEUE = sig
  type elt
  type t
  val emp : t
  val ins : elt * t  $\rightarrow$  t
  val rem : t  $\rightarrow$  (elt * t) option
end

structure QL :> QUEUE = struct
  type elt = bool
  type t = elt list
  val emp = nil
  val ins = cons
  fun rem nil = NONE
    | let val (x,q') = rev q in SOME (x, rev q')
end
```

Modules and Phases

Types such as $QL.\text{elt} \rightarrow QL.\text{t}$ **threaten** the phase distinction!

- Structure QL has static and dynamic components.
- What, then, is type equality?

Moggi addressed this concern **analytically**:

- All modules **decompose** into **static** and **dynamic** parts.
- Maps between modules inherently respect phase separation.

For example, $q:QUEUE \vdash M : QUEUE$ separates into two parts:

- M^{st} defines types elt and t in terms of $q^{st}.\text{elt}$ and $q^{st}.\text{t}$.
- M^{dy} defines values emp , etc in terms of types and values in q .

Sharing Specifications

Coherence is specified by equational **sharing specifications**.

```
functor Layer
```

```
(structure Lower : LAYER and Packet : PACKET
```

```
  sharing Lower.Packet.t = Packet.t
```

Supports composition from **pre-existing** components!

- **Avoids** anticipatory abstraction over shared components.
- **Supports** off-the-shelf re-use.

But what do sharing specifications **mean**?

Types for program modules

Dreyer, Rossberg and Russo: full-scale analytic account in their F'ing Modules.

- Phase separation into System F_ω .
- Rich module structure, including abstraction and sharing.

Here we consider a **synthetic** account.

- Modules come **first**: everything is a module.
- Phases are **isolated** declaratively.
- Type equality is **phase-sensitive**.

ModTT: A Type System for Modules

Following MacQueen, start with dependent types:

- A **universe** of **core language** types and programs.
- Dependent **products** $x:\sigma_1 \times \sigma_2$ for **hierarchy**.
- Dependent **functions** $x:\sigma_1 \rightarrow \sigma_2$ for **parameterization**.

Extend dependent types with

- A **lax** account of abstraction and effects.
- A **modal** account of the phase distinction.
- **Extension** types for sharing [cf cubical type theories].

Polymorphism arises from modules abstracting over the universe.

SML was the first full-scale dependently typed programming language!

Structure of ModTT

Basic judgments:

$\Gamma \vdash \sigma \text{ sig}$	signature
$\Gamma \vdash \sigma \equiv \sigma'$	signature equality
$\Gamma \vdash V : \sigma$	module value
$\Gamma \vdash V \equiv V' : \sigma$	module value equality
$\Gamma \vdash M \div \sigma$	module computation
$\Gamma \vdash M \equiv M' \div \sigma$	module computation equality

Structure of ModTT

Signatures:

$$\frac{}{\Gamma \vdash \text{type } sig}$$

$$\frac{\Gamma \vdash \tau : \text{type}}{\Gamma \vdash \text{val}(\tau) sig}$$

$$\frac{\Gamma \vdash \sigma sig}{\Gamma \vdash \diamond \sigma sig}$$

$$\frac{\Gamma \vdash \sigma_0 sig \quad \Gamma, x:\sigma_0 \vdash \sigma_1 sig}{\Gamma \vdash x : \sigma_0 \times \sigma_1 sig}$$

$$\frac{\Gamma \vdash \sigma_0 sig \quad \Gamma, x:\sigma_0 \vdash \sigma_1 sig}{\Gamma \vdash x : \sigma_0 \rightarrow \sigma_1 sig}$$

Structure of ModTT

Module computations and encapsulation:

$$\frac{\Gamma \vdash M \div \sigma}{\Gamma \vdash \{M\} : \diamond\sigma}$$

$$\frac{\Gamma \vdash V : \sigma}{\Gamma \vdash \text{ret}(V) \div \sigma}$$

$$\frac{\Gamma \vdash V : \diamond\sigma \quad \Gamma, X : \sigma \vdash M \div \sigma'}{\Gamma \vdash X \leftarrow V; M \div \sigma'}$$

Sealing is a *pro forma* effect. For $M : \sigma$,

$$M :> \sigma \triangleq X \leftarrow \{M\}; \text{ret}(X)$$

“Generativity”: multiple bind’s induce **distinct** values with distinct type components.

Modal Formulation of Phases

Propositional signature specifies static phase:

$$\frac{}{\Gamma \vdash \blacksquare_{\text{st}} \text{sig}} \qquad \frac{\Gamma \vdash V, V' : \blacksquare_{\text{st}}}{\Gamma \vdash V \equiv V' : \blacksquare_{\text{st}}}$$

Static equivalence of signatures:

$$\Gamma, \blacksquare_{\text{st}} \vdash \sigma \equiv \sigma'$$

Type checking respects static equivalence:

$$\frac{\Gamma \vdash V : \sigma \quad \Gamma, \blacksquare_{\text{st}} \vdash \sigma \equiv \sigma'}{\Gamma \vdash V : \sigma'}$$

(and similarly for computations)

Static Typing

The static phase **identifies** expressions:

$$\frac{\Gamma \vdash \tau : \text{type} \quad \Gamma \vdash \blacksquare_{\text{st}}}{\Gamma \vdash * : \text{val}(\tau)}$$

$$\frac{\Gamma \vdash e : \text{val}(\tau) \quad \Gamma \vdash \blacksquare_{\text{st}}}{\Gamma \vdash e \equiv * : \text{val}(\tau)}$$

and module computations:

$$\frac{\Gamma \vdash \sigma \text{ sig} \quad \Gamma \vdash \blacksquare_{\text{st}}}{\Gamma \vdash * \div \sigma}$$

$$\frac{\Gamma \vdash M \div \sigma \quad \Gamma \vdash \blacksquare_{\text{st}}}{\Gamma \vdash M \equiv * \div \sigma}$$

Necessary for **static type checking**, a key design parameter.

Static Extent

Sharing is accounted for by **static extent** signatures:

FORMATION

$$\frac{\begin{array}{l} \Gamma \vdash \sigma \text{ sig} \\ \Gamma, \blacksquare_{\text{st}} \vdash V : \sigma \end{array}}{\Gamma \vdash \{\sigma \mid \blacksquare_{\text{st}} \hookrightarrow V\} \text{ sig}}$$

INTRODUCTION

$$\frac{\begin{array}{l} \Gamma \vdash U : \sigma \\ \Gamma, \blacksquare_{\text{st}} \vdash U \equiv V : \sigma \end{array}}{\Gamma \vdash U : \{\sigma \mid \blacksquare_{\text{st}} \hookrightarrow V\}}$$

ELIMINATION

$$\frac{\Gamma \vdash U : \{\sigma \mid \blacksquare_{\text{st}} \hookrightarrow V\}}{\Gamma \vdash U : \sigma \quad \Gamma, \blacksquare_{\text{st}} \vdash U \equiv V : \sigma}$$

SML sharing is encodable in terms of static extent.

Phased Modalities

The static open induces **open** and **closed** modalities:

- **Purely static:** $\circ_{\text{st}}(\sigma) = \blacksquare_{\text{st}} \rightarrow \sigma$.
- **Purely dynamic:** $\bullet_{\text{st}}(\sigma) = \sigma \vee \blacksquare_{\text{st}}$ (pushout of projections)

Thus, modules of the closed signature has trivial static part:

$$\circ_{\text{st}}(\bullet_{\text{st}}(\sigma)) \cong 1$$

Think of a module as **indexed** over its static aspect.

- Static aspect isolates index (type components).
- Dynamic aspect “hides” static aspect by “shifting” it to the dynamic.

Relational Parametricity

Reynolds introduced **parametricity** to explain **data abstraction**.

- Implementors **provide** the type and its implementation.
- Clients are **polymorphic** in the abstract type.

Parametricity theorem: If $e : \forall t. \sigma$, then for all τ, τ' and all $R : \tau \leftrightarrow \tau'$,

$$e[\tau] =_{\sigma} e[\tau'] \quad (\text{rel. } t \mapsto R)$$

Consequently, no client can distinguish **corresponding** implementations of an ADT:

- Define correspondence relation between implementation types.
- Show that the operations preserve this relation.

In short, an application of (binary, heterogeneous) **Tait computability**.

Relational Parametricity

Reynolds worked **analytically** with System F:

- Function types a la Tait: $R_{\tau_1 \rightarrow \tau_2} = R_{\tau_1} \rightarrow R_{\tau_2}$.
- Polymorphic types a la Girard: quantify over **admissible** types.

Extending Reynolds to ModTT poses challenges:

- Universe permits types as **outputs**, not just **inputs**.
- Mixed-phase dependent types.
- Extent types, static equivalence.
- Modality for effects.

Generalize parametricity **relations** to parametricity **structures**.

- **Proof-relevant**, for the universe (classifier of classifiers).
- **Binary, heterogeneous**, as in Reynolds.
- **Phase-separated**, in two senses!

Logical Relations as Types

A **synthetic** account of parametricity, ParamTT.

- All types are parametricity structures.
- Itself a type system for phase-separated modules.
- Phase distinction between **syntax** and **semantics**.

Syntactic phase, $\blacksquare_{\text{syn}}$.

- Purely syntactic: $\circ_{\text{syn}}(\sigma) = \blacksquare_{\text{syn}} \rightarrow \sigma$.
- Purely semantic: $\bullet_{\text{syn}}(\sigma) = \sigma \vee \blacksquare_{\text{syn}}$.

Implicitly **binarized** a la Wadler with **left** and **right** parts:

- $\blacksquare_{\text{syn}} = \blacksquare_{\text{syn}/l} \vee \blacksquare_{\text{syn}/r}$
- $\blacksquare_{\text{syn}/l} \wedge \blacksquare_{\text{syn}/r} = \perp$.

Interpretation of Signatures and Modules

Signatures are interpreted as

- A **syntactic** signature, together with
- A parametricity structure for its elements.

$$\begin{aligned} \text{Sig} &: \{ \mathcal{U} \mid \blacksquare_{\text{syn}} \hookrightarrow \text{Sig} \} \\ &\cong \sigma : \text{Sig} \times \{ \mathcal{U} \mid \blacksquare_{\text{syn}} \hookrightarrow \sigma \} \end{aligned}$$

Correspondingly, modules of a signature extract that structure:

$$\begin{aligned} \text{Mod} &: \{ \text{Sig} \rightarrow \mathcal{U} \mid \blacksquare_{\text{syn}} \hookrightarrow \text{Mod} \} \\ \text{Mod}(\sigma, \sigma^*) &\triangleq \sigma^* \end{aligned}$$

Dependent Functions

For $\sigma_0 : \text{Sig}$ and $\sigma_1 : \sigma_0 \rightarrow \text{Sig}$,

$$\begin{aligned}\sigma_\pi^* &: \{ \mathcal{U} \mid \blacksquare_{\text{syn}} \hookrightarrow \text{Mod}(\sigma_\pi) \} \\ &\cong x : \text{Mod}(\sigma_0) \rightarrow \text{Mod}(\sigma_1(x))\end{aligned}$$

where

$$\sigma_\pi \triangleq x : \text{Mod}(\sigma_0) \rightarrow \sigma_1(x)$$

(the syntactic function signature, under $\blacksquare_{\text{syn}}$)

Provides interpretation of abstraction and application.

Universe of core language types:

$$\begin{aligned} \text{Type} &: \{ \text{Sig} \mid \blacksquare_{\text{syn}} \hookrightarrow \text{Type} \} \\ &\triangleq (\text{Type}, \text{Type}^*) \end{aligned}$$

The semantics of types is given by purely dynamic parametricity structures:

$$\begin{aligned} \text{Type}^* &: \{ \mathcal{U} \mid \blacksquare_{\text{syn}} \hookrightarrow \text{Type} \} \\ &\cong \tau : \text{Type} \times \{ \mathcal{U}_{\bullet_{\text{st}}} \mid \blacksquare_{\text{syn}} \hookrightarrow \text{Val}(\tau) \} \end{aligned}$$

The parametricity structure for values is extracted from that of the universe:

$$\begin{aligned} \text{Val} &: \{ \text{Type} \rightarrow \text{Sig} \mid \blacksquare_{\text{syn}} \hookrightarrow \text{Val} \} \\ \text{Val}(\tau, \tau^*) &\triangleq (\text{Val}(\tau), \tau^*) \end{aligned}$$

Interpretation of Booleans

Booleans are interpreted as purely dynamic, purely semantic structure:

$$\begin{aligned} Bool &: \{ Type \mid \blacksquare_{\text{syn}} \hookrightarrow Bool \} \\ &\triangleq (Bool, Bool^*) \end{aligned}$$

Booleans are observably either true or false:

$$\begin{aligned} Bool^* &: \{ \mathcal{U} \mid \blacksquare_{\text{syn}} \hookrightarrow Val(bool) \} \\ &\cong b : Val(bool) \times \bullet_{\text{syn}} \bullet_{\text{st}} (b^* : 2 \times case(b^*; true; false)) \end{aligned}$$

Could also be given Reynolds-style by isolating the **propositional** parametricity structures (subterminal over each syntactic object).

Justifying ParamTT

Parametricity structures can be explained in terms of toposes:

- ParamTT is the internal language of a pre-sheaf topos.
- Syntactic phase distinction: **glueing syntax to semantics**.
- Static phase distinction: **phase-separated sets**.

All an instance of Sterling's **Synthetic Tait Computability**.

- Synthetic formulation of **proof-relevant logical relations**.
- Normalization for Cartesian cubical type theory.

Further Reading

Details, comparisons, and citations:

Jon Sterling and H., "Logical Relations as Types: Proof-Relevant Parametricity for Program Modules," J. ACM v.68, n.6, October 2022.

Sterling's dissertation:

Jon M. Sterling, "First Steps in Synthetic Tait Computability." CMU SCS Ph.D. Thesis, October 2021.

Cost and Behavior in Type Theory

Dependent type theory expresses **behavior** of programs.

- Insertion sort: $\text{insertsort} : \text{seq} \rightarrow \text{seq}$
- Merge sort: $\text{mergesort} : \text{seq} \rightarrow \text{seq}$

Extensionally, these are **equal**:

$$\text{insertsort} \equiv \text{mergesort} : \text{seq} \rightarrow \text{seq}$$

Yet, they have **different** costs (number of comparisons on size n):

- $\text{insertsort} : \text{seq} \xrightarrow{n^2} \text{seq}$
- $\text{mergesort} : \text{seq} \xrightarrow{n \lg n} \text{seq}$

But equal things cannot have different properties. Phases to the rescue

Integrating Cost and Behavior

CalF = Cost-Aware Logical Framework.

- Extends Pedrot and Tabareau's ∂ CBPV integrating effects and dependency.
- Sole effect: **step counting** for recording resource usage.

Phase μ_{ext} distinguishes **intensional** from **extensional** properties.

- $\text{insertsort} : \text{seq} \xrightarrow{n^2} \text{seq}$
- $\text{mergesort} : \text{seq} \xrightarrow{n \lg n} \text{seq}$
- $\mu_{\text{ext}} \vdash \text{insertsort} \equiv \text{mergesort} : \text{seq} \rightarrow \text{seq}$

Thus, intensional codifies **algorithms**, extensional codifies **functions**.

Calf is equipped with a writer monad for **step counting**

- **step(e)**: increment step-count, then behave as e .
- $\mathbf{\mu}_{\text{ext}} \vdash \text{step}(e) \equiv e : \tau$: disregard resource accounting.

Steps have no **intrinsic** meaning.

- Defined equationally, not via an operational interpretation.
- (Under development: realizability interpretation.)

Resources are **abstract** and **problem-specific**:

- Number of comparisons for `mergesort` and `insertsort`.
- Number of modulus operations for `GCD`.
- Number of queue operations for `batched-queues`.

Open and Closed Modalities

The open modality, $\circ_{\text{ext}}(A)$, isolates **behavior**.

The closed modality, $\bullet_{\text{ext}}(A)$, ensures **non-interference**:

- $\circ_{\text{ext}}(\bullet_{\text{ext}}(\tau)) \cong \mathbf{1}$: no extensional component.
- Consequently, any $\bullet_{\text{ext}}(A) \rightarrow \circ_{\text{ext}}(B)$ is **constant**.

In short **behavior** cannot depend on the step count.

- Counter type is $\bullet_{\text{ext}}(N)$, for the sequential case.
- And is $\bullet_{\text{ext}}(N) \times \bullet_{\text{ext}}(N)$, for the parallel case.

Expressing Cost

Costs have an additive monoidal structure for **work**:

$$\text{step}_X^0(e) = e \qquad \text{step}_X^c(\text{step}_X^d(e)) = \text{step}^{c+d}(e)$$

Costs **commute** with computations:

$$\text{bind}(\text{step}_{F(A)}^c(e); f) = \text{step}_X^c(\text{bind}(e; f))$$

$$\lambda x. \text{step}_X^c(e) = \text{step}_{A \rightarrow X}^c(\lambda x. e)$$

(Parallelism adds multiplicative monoid for **span** a la Blelloch)

Cost Bounds

The cost of a computation $\text{hasCost}(B, e, c)$ is defined as

$$b : B \times (e =_{F(B)} \text{step}^c(\text{ret}(b)))$$

For $c : A \rightarrow N$, the type $a : A \xrightarrow{c} B$ is short for

$$f : (a : A \rightarrow B) \times (a : A \rightarrow \text{hasCost}(B(a), f(a), c(a))).$$

Similarly, $\text{isBounded}(B, e, c)$ specifies an **upper bound**:

$$c' : N \times \circ_{\text{ext}}(c \leq_N c') \times \text{hasCost}(B, e, c')$$

Comparison is in **extensional** mode. Allows for using behavior to analyze cost.

A Nicely Closed World

Calf, as a dependent type theory, is limited to **total** functions.

- Type-specific induction/recursion.
- Awkward compared to general recursion.

How to express efficient algorithms in Calf?

- All algorithms are instrumented with a “clock” for recursive calls.
- If insufficient time is available, terminates with partial result.

The clock is an artificiality in the behavioral setting, but a natural here.

- Typical cost measures **bound** recursion depth.
- Use cost analysis to “set the clock.”

Calf Methodology

- 1 Instrument code for cost accounting,

$$\text{mod}_{instr}(x, y) = \text{step}(x \% y)$$

- 2 Define **clocked** version of algorithm,

$$\text{gcd}_{clocked} : \text{nat} \rightarrow (\text{nat} \times \text{nat}) \rightarrow \text{nat}.$$

$$\lambda(k).\lambda(x, y).\dots \text{gcd}_{clocked}(k - 1)(y, \text{mod}_{instr}(y, \text{mod}_{instr}(x, y)))$$

- 3 Define **cost recurrence** by any means (cost of recurrence is irrelevant)

$$\text{gcd}_{depth}(x, y) : \text{nat} \times \text{nat} \rightarrow \text{nat}$$

- 4 Define complete algorithm:

$$\text{gcd}(x, y) = \text{gcd}_{clocked}(\text{gcd}_{depth}(x, y))(x, y)$$

The recurrence determines an upper bound on modulus operations:

$$\text{isBounded}(N; \text{gcd}(x, y); \text{gcd}_{\text{depth}}(x, y))$$

The depth recurrence can be solved:

$$\text{gcd}_{\text{depth}}(x, y) \leq \text{Fib}^{-1}(x) + 1$$

Combining these,

$$\text{isBounded}(N; \text{gcd}(x, y); \text{Fib}^{-1}(x, y))$$

These, and other (sequential and parallel) bounds, are fully mechanized in Agda.

Further Reading

For background, comparisons, citations, and full development:

Yue Niu, Jon M. Sterling, Harrison Grodin, and H, "A Cost-Aware Logical Framework." To appear, ACM SIGPLAN Symp. on Princ. of Prog. Lang. (POPL). Philadelphia, January 2022 (to appear).

Mechanization in Agda:

<https://github.com/jonsterling/agda-calf>

Watch for Niu's Ph.D., expected 2023!

Phase Distinctions Abound!

Information flow security (ongoing work, with Sterling and Stephanie Balzer).

- **Public** (vs private) phase, \blacksquare_{pub} .
- **Public equivalence**: all private computations are equated.
- Scales naturally to a lattice of levels.

Debugging vs delivery: $\blacksquare_{deliver}$.

- Instrument code with profiling and tracing information (a la step counting).
- Active under debug phase, disregarded under $\blacksquare_{deliver}$.
- Presented at ML Workshop,
<https://www.cs.cmu.edu/~rwh/papers/multiphase/mlw.pdf>

Thank you!

Queue Signature

```
signature QUEUE = sig
  type elt = bool
  type t
  val emp : t
  val ins : elt * t  $\rightarrow$  t
  val rem : t  $\rightarrow$  elt * t
end
```

Queue Implementation: Lists

```
structure QL : QUEUE = struct
  type elt = bool
  type t = elt list
  val emp = nil
  fun ins (x, q) = ret (x :: q)
  fun rem q =
    bind val rev_q ← rev q in
    case rev_q of
    | nil ⇒ throw
    | x :: xs ⇒
      bind val rev_xs ← rev xs in
      ret (f, rev_xs)
end
```

Queue Implementation: Pair of Lists

```
structure QLL : QUEUE = struct
  type elt = bool
  type t = elt list * elt list
  val emp = (nil, nil)
  fun ins (x, (fs, rs)) = ret (fs, x :: rs)
  fun rem (fs, rs) =
    case fs of
    | nil =>
      bind val rev_rs ← rev rs in
      (case rev_rs of
      | nil => throw
      | x::rs' => ret (x, rs', nil))
    | x::fs' => ret (x, fs', rs)
end
```


Correspondence Structure

A simulation over $C = [\mathbf{r}_{\text{syn}/l} \hookrightarrow \text{QL}, \mathbf{r}_{\text{syn}/r} \hookrightarrow \text{QLL}]$ consists of the following data:

$$t : \{\text{Mod}(\text{type}) \mid \mathbf{r}_{\text{syn}} \hookrightarrow \text{QC.t}\}$$

$$\text{emp} : \{\text{Mod}(\langle t \rangle) \mid \mathbf{r}_{\text{syn}} \hookrightarrow \text{QC.emp}\}$$

$$\text{ins} : \{\text{Mod}(\langle \text{bool} * t \rightarrow t \rangle) \mid \mathbf{r}_{\text{syn}} \hookrightarrow \text{QC.ins}\}$$

$$\text{rem} : \{\text{Mod}(\langle t \rightarrow \text{bool} * t \rangle) \mid \mathbf{r}_{\text{syn}} \hookrightarrow \text{QC.rem}\}$$

$$\text{invariant} : \{\mathcal{U}_{\bullet_{\text{st}}}^{\alpha} \mid \mathbf{r}_{\text{syn}} \hookrightarrow \bullet_{\text{st}} \circ_{\text{syn}} \text{Mod}(\text{QC.t})\}$$

$$\text{invariant} \cong \sum_{q: \circ_{\text{syn}} \text{Mod}(\langle \text{QC.t} \rangle)} \bullet_{\text{syn}}(\{\vec{x}, \vec{y}, \vec{z} : \bullet_{\text{st}}(\text{bits}) \mid \vec{x} = (\vec{y} + \text{rev}(\vec{z})) \wedge \dots\})$$

$$\dots = q = [\mathbf{r}_{\text{syn}/l} \hookrightarrow \lceil \vec{x} \rceil \mid \mathbf{r}_{\text{syn}/r} \hookrightarrow (\lceil \vec{y} \rceil, \lceil \vec{z} \rceil)]$$