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

```ml
signature QUEUE = sig
  type elt
  type t
  val emp : t
  val ins : elt * t \to t
  val rem : t \to (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
```
Types such as $\text{QL.elt} \to \text{QL.t}$ 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, $q : \text{QUEUE} \vdash M : \text{QUEUE}$ separates into two parts:

- $M^{st}$ defines types $\text{elt}$ and $\text{t}$ in terms of $q^{st}.\text{elt}$ and $q^{st}.\text{t}$.
- $M^{dy}$ defines values $\text{emp}$, etc in terms of types and values in $q$. 
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 Fω.
- Rich module structure, including abstraction and sharing.

Here we consider a synthetic account.

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

\[ \Gamma \vdash \sigma \text{ sig} \quad \text{signature} \]
\[ \Gamma \vdash \sigma \equiv \sigma' \quad \text{signature equality} \]
\[ \Gamma \vdash V : \sigma \quad \text{module value} \]
\[ \Gamma \vdash V \equiv V' : \sigma \quad \text{module value equality} \]
\[ \Gamma \vdash M \div \sigma \quad \text{module computation} \]
\[ \Gamma \vdash M \equiv M' \div \sigma \quad \text{module computation equality} \]
Structure of ModTT

Signatures:

\[
\begin{align*}
\Gamma \vdash \tau : \text{type} & \quad \Gamma \vdash \sigma \text{ sig} \\
\Gamma \vdash \text{val}(\tau) \text{ sig} & \quad \Gamma \vdash \Diamond \sigma \text{ sig}
\end{align*}
\]

\[
\begin{align*}
\Gamma \vdash \sigma_0 \text{ sig} \quad \Gamma, x : \sigma_0 \vdash \sigma_1 \text{ sig} & \quad \Gamma \vdash \sigma_0 \text{ sig} \quad \Gamma, x : \sigma_0 \vdash \sigma_1 \text{ sig} \\
\Gamma \vdash x : \sigma_0 \times \sigma_1 \text{ sig} & \quad \Gamma \vdash x : \sigma_0 \to \sigma_1 \text{ sig}
\end{align*}
\]
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:

\[ \Gamma \vdash \mathcal{O}_{st} \text{ sig} \]

\[ \Gamma \vdash V, V' : \mathcal{O}_{st} \]

\[ \Gamma \vdash V \equiv V' : \mathcal{O}_{st} \]

Static equivalence of signatures:

\[ \Gamma, \mathcal{O}_{st} \vdash \sigma \equiv \sigma' \]

Type checking respects static equivalence:

\[ \Gamma \vdash V : \sigma \quad \Gamma, \mathcal{O}_{st} \vdash \sigma \equiv \sigma' \]

\[ \Gamma \vdash V : \sigma' \]

(and similarly for computations)
The static phase identifies expressions:

\[
\begin{align*}
\Gamma \vdash \tau : \text{type} & \quad \Gamma \vdash \sigma_{st} \\
\Gamma \vdash * : \text{val}(\tau) &
\end{align*}
\]

and module computations:

\[
\begin{align*}
\Gamma \vdash \sigma \text{ sig} & \quad \Gamma \vdash \sigma_{st} \\
\Gamma \vdash * \div \sigma &
\end{align*}
\]

Necessary for static type checking, a key design parameter.
Sharing is accounted for by *static extent* signatures:

\[
\begin{align*}
&\text{FORMATION} & \quad \Gamma \vdash \sigma \quad \text{sig} \\
& & \quad \Gamma, \sigma_{st} \vdash V : \sigma \\
& & \quad \Gamma \vdash \{\sigma \mid \sigma_{st} \leftrightarrow V\} \quad \text{sig} \\
&\text{INTRODUCTION} & \quad \Gamma \vdash U : \sigma \\
& & \quad \Gamma, \sigma_{st} \vdash U \equiv V : \sigma \\
& & \quad \Gamma \vdash U : \{\sigma \mid \sigma_{st} \leftrightarrow V\} \\
&\text{ELIMINATION} & \quad \Gamma \vdash U : \{\sigma \mid \sigma_{st} \leftrightarrow V\} \\
& & \quad \Gamma \vdash U : \sigma \\
& & \quad \Gamma, \sigma_{st} \vdash U \equiv V : \sigma
\end{align*}
\]

SML sharing is encodable in terms of static extent.
The static open induces open and closed modalities:

- **Purely static**: $\bigcirc_{st}(\sigma) = \mathcal{F}_{st} \to \sigma$.
- **Purely dynamic**: $\bullet_{st}(\sigma) = \sigma \lor \mathcal{F}_{st}$ (pushout of projections)

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

$$\bigcirc_{st}(\bullet_{st}(\sigma)) \cong 1$$

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.
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 $\tau, \tau'$ and all $R : \tau \leftrightarrow \tau'$,

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

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**.
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!
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, \( \mathcal{R}_{\text{syn}} \).

- Purely syntactic: \( \circ_{\text{syn}}(\sigma) = \mathcal{R}_{\text{syn}} \rightarrow \sigma \).
- Purely semantic: \( \bullet_{\text{syn}}(\sigma) = \sigma \lor \mathcal{R}_{\text{syn}} \).

Implicitly binarized a la Wadler with left and right parts:

- \( \mathcal{R}_{\text{syn}} = \mathcal{R}_{\text{syn}}/l \lor \mathcal{R}_{\text{syn}}/r \)
- \( \mathcal{R}_{\text{syn}}/l \land \mathcal{R}_{\text{syn}}/r = \bot \).
Interpretation of Signatures and Modules

Signatures are interpreted as

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

\[
\text{Sig} : \{ \mathcal{U} \mid \Box \text{syn} \leftrightarrow \text{Sig} \} \\
\cong \sigma : \text{Sig} \times \{ \mathcal{U} \mid \Box \text{syn} \leftrightarrow \sigma \}
\]

Correspondingly, modules of a signature extract that structure:

\[
\text{Mod} : \{ \text{Sig} \to \mathcal{U} \mid \Box \text{syn} \leftrightarrow \text{Mod} \}
\]

\[
\text{Mod}(\sigma, \sigma^*) \triangleq \sigma^*
\]
Dependent Functions

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

$$\sigma_\pi^* : \{ \mathcal{U} \mid \square_{\text{syn}} \leftrightarrow \text{Mod}(\sigma_\pi) \}$$

$$\cong x : \text{Mod}(\sigma_0) \rightarrow \text{Mod}(\sigma_1(x))$$

where

$$\sigma_\pi \overset{\triangle}{=} x : \text{Mod}(\sigma_0) \rightarrow \sigma_1(x)$$

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

Provides interpretation of abstraction and application.
Universe of core language types:

\[
\text{Type} : \{ \text{Sig} \mid \text{syn} \leftrightarrow \text{Type} \} \\
\triangleq (\text{Type}, \text{Type}^*)
\]

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

\[
\text{Type}^* : \{ \mathcal{U} \mid \text{syn} \leftrightarrow \text{Type} \} \\
\cong \tau : \text{Type} \times \{ \mathcal{U}_{\text{st}} \mid \text{syn} \leftrightarrow \text{Val}(\tau) \}
\]

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

\[
\text{Val} : \{ \text{Type} \to \text{Sig} \mid \text{syn} \leftrightarrow \text{Val} \} \\
\text{Val}(\tau, \tau^*) \triangleq (\text{Val}(\tau), \tau^*)
\]
Interpretation of Booleans

Booleans are interpreted as purely dynamic, purely semantic structure:

\[
Bool : \{ \text{Type} | \text{syn} \mapsto Bool \}
\]
\[
\triangleq (Bool, Bool^*)
\]

Booleans are observably either true or false:

\[
Bool^* : \{ \forall | \text{syn} \mapsto Val(\text{bool}) \}
\]
\[
\cong b : Val(\text{bool}) \times \text{syn} \bullet \text{st}(b^* : 2 \times \text{case}(b^*; \text{true}; \text{false}))
\]

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


Sterling’s dissertation:

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 . . . .
Calf = Cost-Aware Logical Framework.

- Extends Pedrot and Tabareau’s $\partial$CBPV integrating effects and dependency.
- Sole effect: step counting for recording resource usage.

Phase $\Box_{\text{ext}}$ distinguishes intensional from extensional properties.

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

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`.
- `\text{\texttt{\textbackslash 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 insersort.
- Number of modulus operations for GCD.
- Number of queue operations for batched-queues.
Open and Closed Modalities

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

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

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

In short behavior cannot depend on the step count.

- Counter type is $\Box_{\text{ext}}(\mathcal{N})$, for the sequential case.
- And is $\Box_{\text{ext}}(\mathcal{N}) \times \Box_{\text{ext}}(\mathcal{N})$, for the parallel case.
Costs have an additive monoidal structure for work:

\[ \text{step}^0_X(e) = e \quad \text{step}^c_X(\text{step}^d_X(e)) = \text{step}^{c+d}(e) \]

Costs *commute* with computations:

\[ \text{bind}(\text{step}^c_{F(A)}(e); f) = \text{step}^c_X(\text{bind}(e; f)) \]

\[ \lambda x.\text{step}^c_X(e) = \text{step}^c_{A \rightarrow X}(\lambda x. e) \]

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

The cost of a computation 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 \(a : A \rightarrow B\) is short for

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

Similarly, 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.”
1. Instrument code for cost accounting,
\[
\text{mod}_{\text{instr}}(x, y) = \text{step}(x \% y)
\]

2. Define \textit{clocked} version of algorithm,
\[
\text{gcd}_{\text{clocked}} : \text{nat} \rightarrow (\text{nat} \times \text{nat}) \rightarrow \text{nat}.
\]
\[
\lambda(k).\lambda(x, y) \ldots . \text{gcd}_{\text{clocked}}(k - 1)(y, \text{mod}_{\text{instr}}(y, \text{mod}_{\text{instr}}(x, y)))
\]

3. Define \textit{cost recurrence} by any means (cost of recurrence is irrelevant)
\[
\text{gcd}_{\text{depth}}(x, y) : \text{nat} \times \text{nat} \rightarrow \text{nat}
\]

4. Define complete algorithm:
\[
\text{gcd}(x, y) = \text{gcd}_{\text{clocked}}(\text{gcd}_{\text{depth}}(x, y))(x, y)
\]
The recurrence determines an upper bound on modulus operations:

\[ \text{isBounded}(N; \gcd(x, y); \gcd_{\text{depth}}(x, y)) \]

The depth recurrence can be solved:

\[ \gcd_{\text{depth}}(x, y) \leq \text{Fib}^{-1}(x) + 1 \]

Combining these,

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

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

For background, comparisons, citations, and full development:

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, $\mathcal{P}_{pub}$.
- **Public equivalence**: all private computations are equated.
- Scales naturally to a lattice of levels.

Debugging vs delivery: $\mathcal{P}_{deliver}$.

- Instrument code with profiling and tracing information (a la step counting).
- Active under debug phase, disregarded under $\mathcal{P}_{deliver}$.
- Presented at ML Workshop, 
Thank you!
signature QUEUE = sig
  type elt = bool
  type t
  val emp : t
  val ins : elt * t ↴ t
  val rem : t ↦ 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 $C = [\text{\textbf{syn}}/l \leftrightarrow \text{QL}, \text{\textbf{syn}}/r \leftrightarrow \text{QLL}]$ consists of the following data:

- $t : \{\text{Mod(\text{type})} | \text{\textbf{syn} } \leftrightarrow \text{QC.t}\}$
- $emp : \{\text{Mod(\{t\})} | \text{\textbf{syn} } \leftrightarrow \text{QC.emp}\}$
- $ins : \{\text{Mod(\{bool \ast t \rightarrow t\})} | \text{\textbf{syn} } \leftrightarrow \text{QC.ins}\}$
- $rem : \{\text{Mod(\{t \rightarrow bool \ast t\})} | \text{\textbf{syn} } \leftrightarrow \text{QC.rem}\}$

\[
\text{invariant} : \{\text{\textbf{U}} \overset{\text{\textbf{st}}}{\alpha} | \text{\textbf{syn} } \leftrightarrow \text{\textbf{st}} \circ \text{\textbf{syn}} \text{Mod(QC.t)}\}
\]

\[
\text{invariant} \cong \sum_{q : \text{\textbf{syn}} \text{Mod(\{QC.t\})}} \text{\textbf{syn}}(\{\vec{x}, \vec{y}, \vec{z} : \text{\textbf{st}}(\text{bits}) | \vec{x} = (\vec{y} + \text{\textbf{rev}}(\vec{z})) \land \ldots \})
\]

\[
\ldots = q = [\text{\textbf{syn}}/l \leftrightarrow [\vec{x}] | \text{\textbf{syn}}/r \leftrightarrow ([\vec{y}], [\vec{z}])]
\]