

Shallow Types for Insightful Programs

Grace is Optional, Performance is Not

Richard Roberts

School of Engineering and Computer Science
Victoria University of Wellington
rykardo.r@gmail.com

Michael Homer

School of Engineering and Computer Science
Victoria University of Wellington
mwh@ecs.vuw.ac.nz

Stefan Marr

School of Computing
University of Kent
s.marr@kent.ac.uk

James Noble

School of Engineering and Computer Science
Victoria University of Wellington
kjax@ecs.vuw.ac.nz

Abstract

Languages with explicit dynamic type checking are increasing in popularity in both practical development and programming education. Unfortunately, current implementations of these languages perform worse than either purely statically or purely dynamically typed languages. We show how virtual machines can use common optimizations to remove redundancy in dynamic type checking, by adding shallow structural type checks to Moth, a Truffle-based interpreter for Grace. Moth runs programs with dynamic type checks roughly as fast as programs without checks, so developers do not need to disable checks in production code, and educators can teach types without also teaching that types slow programs down.

CCS Concepts • Software and its engineering → Just-in-time compilers; Object oriented languages; Interpreters;

Keywords dynamic type checking, gradual types, optional types, Grace, Moth, object-oriented programming

ACM Reference Format:

Richard Roberts, Stefan Marr, Michael Homer, and James Noble. 1997. Shallow Types for Insightful Programs: Grace is Optional, Performance is Not. In *Proceedings of ACM Woodstock conference (WOODSTOCK'97)*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.475/123_4

1 Introduction

Dynamic languages are increasingly prominent in the software industry. Building on the pioneering work of Self [19], much work in academia and industry has gone into making

them more efficient [11, 12, 21–23, 52]. Just-in-time compilers have taken JavaScript, for example, from a naïvely interpreted language barely suitable for browser scripting, to a highly efficient ecosystem that is sweeping across industry and academia [55].

With these performance gains, dynamic languages are used to build larger and larger systems, which leads to typing approaches being adopted to support programmer productivity and document a program's structures. Two important approaches are optional [14] and gradual typing [38, 39]. These are applied to dynamic languages to reap the benefits of typing, but unfortunately also have limitations. With optional or pluggable approaches such as TypeScript [6, 14] types are erased before the execution, limiting the benefit of types to the statically typed parts of programs. In contrast, gradual type systems retain types until run time, performing the checks dynamically, and can give detailed information about type violations via blame tracking [39, 50]. Unfortunately, these gradual systems currently impose significant run-time overheads [5, 26, 35, 36, 44, 45, 49].

We are working on Grace [7], a dynamic language in the tradition of Smalltalk [25], Self [19], and JavaScript that is meant for use in education [18]. While Grace is a dynamic language at its core, we want to have the option to teach students about types, and so Grace supports type annotations which may be checked either statically or dynamically to give students feedback on whether their type annotations are correct. We do not want students to remove types, however, if they discover that types induce a run-time overhead.

Additionally, we are currently maintaining three different implementations to support a variety of educational settings (web browsers, .NET, and JVM), which means a typing approach for Grace ideally requires only small changes to keep these implementations as consistent as possible.

In this paper we illustrate that using an optimizing virtual machine allows dynamic checks of shallow structural types with low overhead and relatively low implementation effort. These checks are inserted naïvely based on local annotations

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WOODSTOCK'97, July 1997, El Paso, Texas USA

© 2016 Copyright held by the owner/author(s).

ACM ISBN 123-4567-24-567/08/06.

https://doi.org/10.475/123_4

and checked eagerly when control flow reaches them: whenever an annotated method is called or an annotated variable is accessed we check types dynamically and terminate the program with a type error if the check fails. Despite this simplistic approach, a just-in-time compiler can eliminate the redundant checks—removing almost all of the checking overhead, resulting in a performance profile aligned with untyped code.

We evaluate this approach with Moth, a Grace implementation on top of Truffle and the Graal just-in-time compiler [52, 53]. Inspired by Richards et al. [36] and Bauman et al. [5], our implementation conflates types with information about the dynamic object structure (maps [19] or object shapes [51]), which allows the just-in-time compiler to reduce redundancy between checking structure and checking types; and consequently, most of the overhead that results from type checking is eliminated.

The contributions of this paper are:

- demonstrating that VM optimisations enable dynamic checks of shallow structural types with low performance cost
- an implementation approach that requires only small changes to existing abstract-syntax-tree interpreters
- an evaluation based on classic benchmarks and benchmarks from the literature on gradual typing

2 Background

This section details our motivation and discusses the technical background for our implementation.

2.1 The Grace Programming Language

We are designing Grace, an object-oriented, imperative, educational programming language, with a focus on introductory programming courses, but also for more advanced study and research [7, 18]. While Grace's syntax draws from the so-called "curly bracket" tradition of C, Java, and JavaScript (with a side order of Pascal) the structure of the language is in many ways closer to Smalltalk (thus Self and Ruby): all computation is via dynamically dispatched "method requests" where the object receiving the request decides which code to run; method names have multiple parts; blocks (lambdas) are used for control flow; and returns within lambdas are "non-local", returning to the method activation in which the block is instantiated [25]. In other ways, Grace is closer to JavaScript than Smalltalk: Grace objects can be created from object literals, rather than by instantiating classes [9, 29] and objects and classes can be deeply nested within each other [31].

Critically, Grace's declarations and methods' arguments and results can be annotated with types, and those types can be checked either statically or dynamically. This means the type system is optional or "pluggable" [14] (removing explicit type annotations should not affect the semantics of a

correct program [39]) and gradual (the type system includes a distinguished "Unknown" type, which matches any other type and is the implicit type for untyped program parts.).

As an educational language [8], absolute performance of an implementation is less important than the performance profile—the way language features affect performance. Increasing absolute performance by several orders of magnitude could let students run larger examples—analyzing billions rather than millions of data points, wayfinding within a city rather than a village, or raytracing higher resolution images a little quicker. On the other hand, issues with a language's performance profile could mean the students will "learn the wrong things". If e.g. a language's built-in cons lists were faster than arrays or hash-tables, students cannot learn the performance benefits of more complex data structures. In the case of Grace, the existing implementations that use type information have the unfortunate property that adding type declarations to a program makes that program run slower—teaching students that *removing* type declarations is an effective optimization technique. Furthermore, this property is shared by other optionally typed languages including Dart's checked mode [15], Reticulated Python [26, 48], and SafeTypeScript [36].

2.2 Moth: Grace on Graal and Truffle

Implementing dynamic languages as state-of-the-art virtual machines can require enormous engineering efforts. Meta-compilation approaches [34] such as RPython [10, 12] and GraalVM [52, 53] reduce the necessary work dramatically, because they allow language implementers to leverage existing VMs and their support for just-in-time compilation and garbage collection.

To leverage this infrastructure, we developed an interpreter for Grace, called Moth [37], by adapting SOMNs [32]. SOMNs is a Newspeak implementation [17] on top of the Truffle framework and the Graal just-in-time compiler, which are part of the GraalVM project. Since Newspeak and Grace are related languages, SOMNs provides a good foundation for a new Grace implementation, allowing us to reach the performance of V8, Google's JavaScript implementation (cf. section 4.2 and Marr et al. [33]) with only moderate effort. SOMNs was changed to parse Grace code, and SOMNs' self-optimizing abstract-syntax-tree nodes were only slightly adapted to conform to Grace's semantics. As a result, Moth is mostly compliant to the Grace specification.

3 Dynamic Type Checks in Grace

The core of Grace's static type system is well described elsewhere [29]; here we explain how these types can be understood dynamically, from a student's or a programmer's point of view. Following from the design goals behind Grace, our motivation for this work is to provide a flexible system to check consistency between an execution of a program and

its type annotations, without significant impact on run-time performance. A secondary goal is to have a design that can be implemented with only a small set of changes to facilitate integration in existing systems.

These goals are shared with much of the other work on gradual type systems, but our context leads to some different choices. First, so that students can see concrete examples of type errors, they should be able to run their programs even if those programs are not type-correct—i.e. Grace’s static type checking is optional, and so an implementation cannot depend on the correctness of a program’s type annotations. Second, while checking Grace’s type annotations statically may be optional, checking them dynamically should not be: any value that flows into a variable, argument, or result annotated with a type must conform to that type annotation. Third, adding type annotations should not degrade a program’s performance, or rather, students should not be encouraged to improve performance by removing type annotations. Unfortunately, existing gradual type implementations do not meet these goals, particularly regarding performance; hence the ongoing debate about whether gradual typing is alive, dead, or some state inbetween [5, 26, 35, 36, 45, 49].

3.1 Design

Our type checks for Grace are designed to be simple, straightforward, and easy for students to understand:

- types are shallow interfaces
- optional type annotations are checked at run time
- failing run-time type checks terminate execution

We illustrate how the type checks work in practice in the context of an exercise where a student is developing a program to record information about vehicles. Grace is structurally typed [7]: an object implements a type whenever it implements all the methods required by a type (as in Go) rather than requiring types to be declared explicitly. A type expresses the requests an object can respond to, for example whether a particular accessor is available, rather than a location in an explicit inheritance hierarchy.

For example, our student can begin developing their vehicle application by defining an object intended to represent a car (listing 1, line 1) and write a method that, given the car object, prints out its registration number (line 5).

```
1 def car = object {
2   var registration is public := "J03553"
3 }
4
5 method printRegistration(v) {
6   print "Registration: {v.registration}"
7 }
```

Listing 1. The start of a simple program for tracking vehicle information.

Next, the student could decide to ensure that any object passed to the `printRegistration` method will respond to the registration request. To get this support, the student first defines the structural type `Vehicle` [46] naming just that method (listing 2, line 1), and then annotates the `printRegistration` method’s argument with that type (listing 2, line 5). This ensures the student will be alerted as soon as a value that does not conform to this expected type is passed into the `printRegistration` method, rather than crashing somewhere in the middle of the implementation of the `print` method.

```
1 type Vehicle = interface {
2   registration
3 }
4
5 method printRegistration(v: Vehicle) {
6   print "Registration: {v.registration}"
7 }
```

Listing 2. Adding a type annotation to a method parameter.

While Grace’s static type system supports full static type checking [7], Grace’s specification requires dynamic type tests to be *shallow*, that is, they check only for the presence of methods in an object, rather than also checking conformance of argument and result types. This is to ensure that the presence or absence of type annotations does not affect the execution of a program, for the reason originally outlined by Boyland [13], thus maintaining a version of the gradual guarantee. The resulting semantics are more-or-less equivalent to type-tag soundness [26]—the difference being that where type-tag soundness supports shallow *nominal* type checks, we support shallow *structural* type checks.

For example, in listing 3, the student creates two cars objects (lines 9 and 17), that conform to an expanded `Vehicle` type (line 1). Note that each version of the `registerTo` method declares a different type for its parameter (lines 12 and 20). When the student runs this program, both `personalCar` and `governmentCar` can be assigned to `Vehicle` because that check considers only that the vehicle has a `registerTo` method, but not the required argument type of that method. At line 25 the student can attempt to register a government car to a person: only when the method is invoked (line 20) the dynamic type test on the argument will fail (the object that is passed in is not a `Department`) even though the body of the `registerTo` method does not rely on the code method that the `Department` annotation requires of the argument.

3.2 Implementation

We have implemented shallow dynamic structural type checks by extending the Moth abstract-syntax-tree (AST) interpreter

```

1 type Vehicle = interface = {
2   registration
3   registerTo(_)
4 }
5
6 type Person = interface { name }
7 type Department = interface { code }
8
9 var personalCar : Vehicle :=
10   object {
11     var registration is public := "DLS018"
12     method registerTo(p: Person) {
13       print "{p.name} registers {self}"
14     }
15   }
16
17 var governmentCar : Vehicle :=
18   object {
19     var registration is public := "FKD218"
20     method registerTo(d: Department) {
21       print "some department {self}"
22     }
23   }
24
25 governmentCar.registerTo(
26   object {
27     var name is public := "Richard"
28   }
29 )

```

Listing 3. A program in development with inconsistent types.

for Grace (section 2.2). Our approach needs to check types of values at run-time:

- the values of arguments are checked after a method is requested, but before the body of the message is executed,
- the value returned by a method is checked after its body is executed, and
- the values of variables are checked whenever written or read by user code.

One of the goals for our approach to dynamic typing was to keep the necessary changes to an existing implementation small, while enabling optimization in highly efficient language runtimes. In an AST interpreter, we can implement this approach by attaching the checks to the relevant AST nodes: the expected types for the argument and return values can be included with the node for requesting a method, and the expected type for a variable can be attached to the nodes for reading from and writing to that variable. In practice, we encapsulate the logic of the check within a new type-checking AST node. Moth's front end was adapted to parse

```

1 class Type:
2   def init(members):
3     self._members = members
4
5   def is_satisfied_by(other: Type):
6     for m in self._members:
7       if not in other._members:
8         return False
9     return True
10
11   def check(obj: Object):
12     t = get_type(obj)
13     return self.is_satisfied_by(t)

```

Listing 4. Sketch of a Type in our system and its check() semantics.

and record type annotations and attach instances of this checking node as children of the existing method, variable read, and variable write nodes.

The check node uses the internal representation of a Grace type (cf. listing 4, line 13) to test whether an observed object conforms to that type. An object satisfies a type if all members required by the type are provided by the object (line 5).

3.3 Optimization

There are two aspects to our implementation that are critical for a minimal overhead solution:

- specialized executions of the type checking node, along with guards to protect these specialized version, and
- a matrix to cache sub-typing relationships to eliminate redundant executions.

The first performance-critical aspect to our implementation is the optimization of the type checking node. We rely on Truffle and its TruffleDSL [27]. This means we provide a number of special cases, which are selected during execution based on the observed concrete kinds of objects. A sketch of our type checking node using a pseudo-code version of the DSL is given in listing 5. A simple optimization is for well known types such as numbers (line 7) or strings (line 11). The methods annotated with @Spec (shorthand for @Specialization) correspond to possible states in a state machine that is generated by the TruffleDSL. Thus, if a check node observes a number or a string, it will check on the first execution only that the expected type, i.e., the one defined by some type annotation, is satisfied by the object by using a static_guard. If this is the case, the DSL will activate this state. For just-in-time compilation, only the activated states and their normal guards are considered. A static_guard is not included in the optimized code. If a check fails, or no


```

441
442 1 class TypeCheckNode(Node):
443 2
444 3     expected: Type
445 4     record: Matrix
446 5
447 6     @Spec(static_guard=expected.check(obj))
448 7     def check(obj: Number):
449 8         pass
450 9
451 10    @Spec(static_guard=expected.check(obj))
452 11    def check(obj: String):
453 12        pass
454 13
455 14    ...
456 15
457 16    @Spec(
458 17        guard=obj.shape==cached_shape,
459 18        static_guard=expected.check(obj))
460 19    def check(obj: Object, @Cached(obj.shape)
461 20        cached_shape: Shape):
462 21        pass
463 22
464 23    @Fallback
465 24    def check_generic(obj: Any):
466 25        T = get_type(obj)
467 26
468 27        if record[T, expected] is unknown:
469 28            record[T, expected] =
470 29                T.is_subtype_of(expected)
471 30
472 31        if not record[T, expected]:
473 32            raise TypeError(
474 33                "{obj} doesn't implement {expected}")

```

Listing 5. An illustration of the type checking node that support type checking

specialization matches, the fallback, i.e., `check_generic` is selected (line 23), which may raise a type error.

For generic objects, we rely on the specialization on line 19, which checks that the object satisfies the expected type. If that is the case, it reads the shape of the object at specialization time, and caches it for later comparisons. Thus, during normal execution, we only need to read the shape of the object and then compare it to the cached one with a simple reference comparison. If the shapes are the same, we can assume the type check passed successfully.

The other performance-critical aspect to our implementation is the use of a matrix to cache sub-typing relationships. The matrix compares types against types, featuring all known types along the columns and the same types again along the rows. A cell in the table corresponds to a sub-typing relationship: does the type corresponding to the row implement the type corresponding to the column? All cells in

the matrix begin as unknown and, as encountered in checks during execution, we populate the table. If a particular relationship has been computed before we can skip the check and instead recall the previously-computed value (line 28). Using this table we are able to eliminate the redundancy of evaluating the same type to type relationships across different checks in the program. To reduce redundancy further we also unify types in a similar way to Java's string interning; during the construction of a type we first check to see if the same set of members is expressed by a previously-created type and, if so, we avoid creating the new instance and provide the existing one instead.

Together the self-specializing type check node and the matrix-based record ensure that our implementation eliminates redundancy, and consequently, we are able to minimize the run-time overhead of our system.

4 Evaluation

To evaluate our approach to dynamic type checking, we first establish the baseline performance of Moth compared to Java and JavaScript, and then assess the impact of the type checks themselves.

4.1 Methodology and Setup

To account for the complex warmup behavior of modern systems [3], we run each benchmark for 1000 iterations in the same VM invocation.¹ Afterwards, we inspected the run-time plots over the iterations and manually determined a cutoff of 350 iterations for warmup, i.e., we discard iterations with signs of compilation. In this work, we do not consider startup performance, because we want to assess the impact of dynamic type checks on the best possible performance. All reported averages use the geometric mean since they aggregate ratios.

All experiments were executed on a machine running Ubuntu Linux 16.04.4, with Kernel 3.13. The machine has two Intel Xeon E5-2620 v3 2.40GHz, with 6 cores each, for a total of 24 hyperthreads. We used ReBench 0.9.1, Java 1.8.0_171, Graal 0.33 (a13b888), Node.js 10.4, and Higgs from 9 May 2018 (aa95240). Our experimental setup is available online to enable reproductions.²

4.2 Are We Fast Yet?

To establish the performance of Moth, we compare it to Java and JavaScript. For JavaScript we chose two implementations, Node.js with V8 as well as the Higgs VM. The Higgs VM is an interesting point of comparison, because Richards et al. [36] used it in their study.

¹ For the Higgs VM, we only use 100 iterations, because of its lower performance. This is sufficient since its compilation approach induces less variation and leads to more stable measurements.

² <https://gitlab.com/richard-roberts/moth-benchmarks/tree/dev>

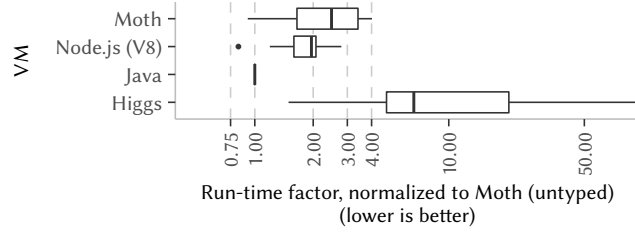


Figure 1. Comparison of Java 1.8, Node.js 10.4, Higgs VM, and Moth. The boxplot depicts the peak-performance results for the Are We Fast Yet benchmarks, each benchmark normalized based on the result for Java. For these benchmarks, Moth is within the performance range of JavaScript, as implemented by Node.js, which makes Moth an acceptable platform for our experiments.

We compare across languages based on the Are We Fast Yet benchmarks [33], which are designed to enable a comparison of the effectiveness of compilers across different languages. To this end, they use only a common set of core language elements. While this reduces the performance relevant differences between languages, the set of core language elements covers only common object-oriented language features with first-class functions. Consequently, these benchmarks are not necessarily a predictor for application performance, but can give a good indication for basic mechanisms such as type checking. Furthermore, in an educational setting, we assume that students will focus on using these basic language features as they learn a new language.

Figure 1 shows the results. We use Java as baseline since it is the fastest language implementation in this experiment. We see that Node.js (V8) is about 1.8x (min. 0.8x, max. 2.8x) slower than Java. Moth is about 2.3x (min. 0.9x, max. 4x) slower than Java. As such, its on average 24% (min. -23%, max. 2.2x) slower than Node.js. Compared to the Higgs VM, which is on these benchmarks 10.7x (min. 1.5x, max. 174x) slower than Java, Moth reaches the performance of Node.js more closely. With these results, we argue that Moth is a suitable platform to assess the impact of our approach to dynamic type checking, because its performance is close enough to state-of-the-art VMs, and run-time overhead is not hidden by slow baseline performance.

4.3 Performance of Dynamic Type Checking

The performance overhead of our dynamic type checking system is assessed based on the Are We Fast Yet benchmarks as well as benchmarks from the gradual-typing literature. The goal was to complement our benchmarks with additional ones that are used for similar experiments and can be ported to Grace. To this end, we surveyed a number of papers [5, 26, 35, 36, 44, 45, 49] and selected benchmarks that have been used by multiple papers. Some of these benchmarks overlapped with the Are We Fast Yet suite, or were

Table 1. Benchmarks selected from literature.

Fannkuch	[26, 49]	
Float	[26, 35, 49]	
Go	[26, 35, 49]	
NBody	[26, 30, 49]	used [33]
Queens	[26, 35, 49]	used [33]
PyStone	[26, 35, 49]	
Sieve	[5, 35, 36, 45]	used [33]
Snake	[5, 35, 36, 45]	
SpectralNorm	[26, 35, 49]	

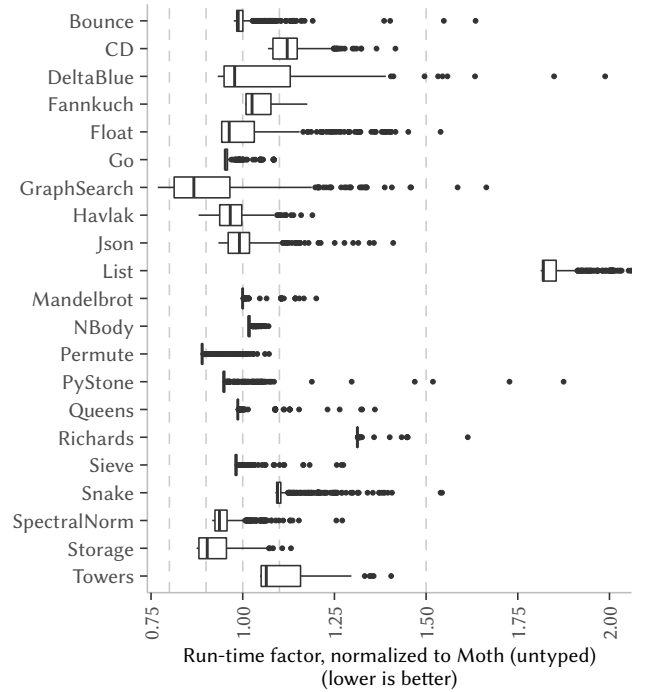


Figure 2. A boxplot comparing the performance of Moth without type checking to Moth with type checking. The plot depicts the run-time overhead on peak performance over the untyped performance. On average, dynamic type checking introduces an overhead of 5% (min. -10%, max. 86%).

available in different versions. While not always behaviorally equivalent, we chose the Are We Fast Yet versions since we already used them to establish the performance baseline. The list of selected benchmarks is given in table 1.

The benchmarks were modified to have complete type information. We modified Moth to report absent type information and ensured it was complete. To assess the performance overhead of type checking, we compare the execution of Moth with all checks disabled, i.e., the baseline version from section 4.2, against an execution that has all checks enabled.

The results are depicted in fig. 2. Overall, we see a peak-performance overhead of 5% (min. -10%, max. 86%).

The benchmark with the highest overhead of 86% is List, which traverses a linked list and has to check the list elements individually in a way that introduces checks that do not coincide with any shape checks on the relevant objects that are performed in the unchecked version. We consider this benchmark a pathological case and discuss it in detail in section 5.

Beside List, the highest overheads are on Richards (31%), CD (13%), Snake (12%), and Towers (10%). Richards has one major component, also a linked list traversal, similar to List. Snake and Towers primarily access arrays in a way that introduces checks that do not coincide with behavior in the unchecked version.

However, in some benchmarks the run time decreased; notably Permute (-10%), GraphSearch (-9%), and Storage (-8%). Permute simply creates the permutations of an array. GraphSearch implements a page rank algorithm and thus is primarily graph traversal. Storage stresses the garbage collector by constructing a tree of arrays. For these benchmarks the introduced checks seem to coincide with shape-check operations already performed in the untyped version. The performance improvement is possibly caused by having checks earlier, which enables the compiler to more aggressively move them out of loops. Another reason could simply be that the extra checks shift the boundaries of compilation units. In such cases, checks might not be eliminated completely, but the shifted boundary between compilation units might mean that objects do not need to be materialized and thus do not need to be allocated, or simply that the generated native code interacts better with the instruction cache of the processor.

While we did not focus on the warmup performance, and are mainly interested in peak performance, fig. 3 shows the first 100 iterations for each benchmark. The run time factor is the result for the typed version over the untyped one. Thus, any increase indicates a slow down because of typing. The gray line in the background indicates a smoothed version of the curve, also indicating a confidence interval.

Focusing on the first iteration, where we assume that most code is executed in the interpreter, the overhead appears minimal. Only the Mandelbrot benchmark shows a noticeable slowdown. However, benchmarks such as Float, PyStone, and GraphSearch show various spikes. Since spikes appear in both directions (speedups and slowdowns), we assume that this merely indicates a shift for instance of garbage collection pauses. This can happen because of different heap configurations possibly triggered by the additional data structures for type information.

4.4 Changes to Moth

Outlined earlier in section 3, a secondary goal of our design was to enable the implementation of our approach to be realized with few changes to the underlying interpreter. This helps to ensure that each Grace implementation can provide type checking in a uniform way.

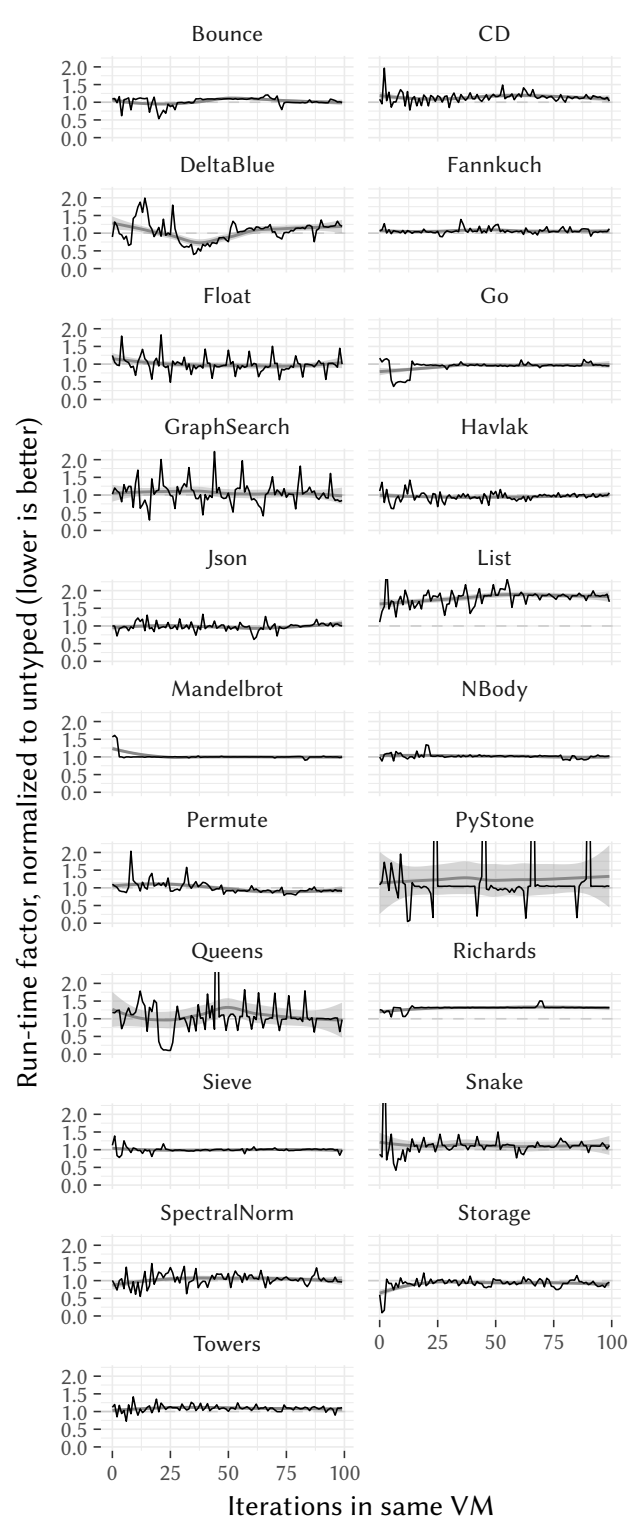


Figure 3. Plot of the run time for the first 100 iterations. The gray line is smoothed version, including a confidence interval, to indicate the trend. The first iteration, i.e., mostly interpreted, seems to be affected significantly only for Mandelbrot.

```

1 var elem: ListElement := headOfList
2 while (...) do {
3   elem := elem.next
4 }

```

Listing 6. Example for dynamic type checks not corresponding to existing checks.

By examining the history of changes maintained by our version control, we estimate that our implementation for Moth required 549 new lines and 59 changes to existing lines. The changes correspond to the implementation of new modules for the type class (179 lines) and the self-specializing type checking node (139 lines), modifications to the front end to extract typing information (115 new lines, 14 lines changes) and finally the new fields and amended constructors for AST nodes (116 new lines, 45 lines changes).

5 Discussion

The VM Could Not Already Know That. One of the key optimizations for our work and the work of others [5, 36] is the use of object shapes to encode information about types (in our case), or type casts and assumptions (in the case of gradually typed systems).

The general idea is that a VM will already use object shapes for method dispatches, field accesses, and other operations on objects. Thus any further use to also imply type information can often be optimized away when the compiler sees that the same checks are done (and therefore can be combined). This is similar to the elimination of other side-effect-free common subexpressions.

This assumption breaks, however, when checks are introduced that do not correspond to those that exist already. As described in section 3, our approach introduces checks for reading and writing to variables. Listing 6 gives an example of a pathological case. It is a loop traversing a linked list. For this example our approach introduces a check, for the `ListElement` type, when (1) assigning to and reading from `elem` and (2) when activating the next method. The checks for reading from `elem` and activating the method can be combined with the dispatch's check on object shape. Unfortunately, the compiler cannot remove the check when writing to `elem`, because it has no information about what value will be returned from `next`, and so it needs to preserve the check to be able to trigger an error on the assignment. For our List benchmark, this check induces an overhead of 86%.

Optimizations As a simplification, we currently check variable access on both reads and writes. This approach simplifies the implementation, because we do not need to adapt all built-ins, i.e., all primitive operations provided by the

interpreter. One optimization could be to avoid read checks. A type violation can normally only occur when writing to a variable, but not when reading. However, to maintain the semantics, this would require us to adapt many primitives; such as operations that activate blocks to check their arguments, or that write to variables or fields. With our current implementation we get errors as soon as user code accesses fields, which simplifies the implementation.

Another optimization could be to use Truffle's approach to self-specialization [54] and propagate type information to avoid redundant checks. At the moment, Truffle interpreters typically use self-specialization to specialize the AST to avoid boxing of primitive types. This is done by speculating that some subtree always returns the expected type. If this is not the case, the return value of the subtree is going to be propagated via an exception, which is caught and triggers respecialization. This idea could possibly be used to encode higher-level type information for return values, too. This could be used to remove redundant checks in the interpreter by simply discovering at run time that whole subexpressions conform to the type annotations.

6 Related Work

Although syntaxes for type annotations in dynamic languages go back at least as far as Lisp [43], the first attempts at adding a comprehensive static type system to a dynamically typed language involved Smalltalk [28], with the first practical system being Bracha's Strongtalk [16]. Strongtalk (independently replicated for Ruby [24]) provided a powerful and flexible static type system, where crucially, the system was *optional* (also known as pluggable [14]). Programmers could run the static checker over their Smalltalk code (or not); either way the type annotations had no impact whatsoever of the semantics of the underlying Smalltalk program.

Siek and Taha [38] introduced the term "gradual typing" to describe the logical extension of this scheme: a dynamic language with type annotations that could, if necessary, be checked at runtime. Siek and Taha build on earlier complementary work extending fully statically typed languages with a "DYNAMIC" type—Abadi et al. [1] is an important early attempt and also surveys previous work. Revived practical adoption of dynamic languages generated revived research interest, leading to the formulation of the "gradual guarantee" [38, 39] to characterize sound gradual type systems: removing type annotations should not change the semantics of a correct program, drawing on Boyland's critical insight that, of course, such a guarantee must by its nature forbid code that can depend on the presence or absence of type declarations elsewhere in the program [13].

Type errors in gradual, or other dynamically checked, type systems will often be triggered by the type declarations, but often those declarations will not be at fault—indeed in a

correctly typed program in a sound gradually typed system, the declarations cannot be at fault because they will have passed the static type checker. Rather, the underlying fault must be somewhere within the barbarian dynamically typed code *trans vallum*. Blame tracking [2, 41, 50] localizes these faults by identifying the point in the program where the system makes an assumption about dynamically typed objects, so can identify the root cause should the assumption fail. Different semantics for blame detect these faults slightly differently, and impose more or less implementation overhead [40, 48, 49].

As with language designs, there seem to be two main implementation strategies for languages mixing dynamic and static type checks: either adding static checks into a dynamic language implementation, or adding support for dynamic types to an implementation that depends on static types for efficiently. Racket, for example, optimizes code with a combination of type inference and type declarations—the Racket IDE “optimizer coach” goes as far as to suggest to programmers type annotations that may improve their program’s performance [42]. In these implementations, values flowing from dynamically to statically typed code must be checked at the boundary. Fully statically typed code needs no dynamic type checks, and so generally performs better than dynamically typed code. Adopting a gradual type system such as Typed Racket [47] allows programmers to explicitly declare types that can be checked statically, removing unnecessary overhead.

On the other hand, systems such as Reticulated Python [48], SafeTypeScript [36], and our work here, takes the opposite approach: adding type checking into just-in-time compiling virtual machines that do not rely on static type declarations. These systems do not use information from type declarations to optimize execution speed, rather the necessity to perform (potentially repeated) dynamic type checks tends to slow programs down, so here code with no type annotations generally performs better than statically typed code, or rather, code with many type annotations. In the limit, these kinds of systems may only ever check types dynamically and may not involve a static type checker at all.

As these systems have come to wider attention, the question of their implementation overheads has become more prominent. Takikawa et al. [45] asked “is sound gradual typing dead?” based on a systematic performance measurement on Typed Racket. The key here is their evaluation methodology, where they constructed a number of different permutations of typed and untyped code, and evaluated performance along the spectrum. Bauman et al. [5] replied to Takikawa et al.’s study, but using Pycket [4], a tracing JIT for Racket, rather than the standard Racket VM, although maintaining full gradually typed Racket semantics. Bauman et al. are able to demonstrate most benchmarks with a slowdown of 2x on average over all configurations. Note that this is not directly comparable to our system, since typed modules do

not need to do any checks at run time. Typed Racket only needs to perform checks at boundaries between typed and untyped modules, however, they use the same essential optimization technique that we apply, using object shapes to encode information about gradual types.

Muehlboeck and Tate [35] also replied to Takikawa et al., using a similar benchmarking methodology applied to Nom, a language with features designed to make gradual types easier to optimize, demonstrating speedups as more type information is added to programs. Their approach enables such type-driven optimizations, but relies on a static analysis which can utilize the type information.

Most recently, Kuhlenschmidt et al. [30] employ an ahead of time (i.e. traditional, static) compiler for a custom language called Grift and demonstrate good performance for code where more than half of the program is annotated with types, and reasonable performance for code without type annotations.

Perhaps the closest to our approach are Greenman and Migeed [26] and Richards et al. [36]. Greenman and Migeed describe dynamically checking “tag-type” soundness for Reticulated Python. As with our work, Greenman and Migeed check only the “top-level” type of an object against a declaration, although Greenman and Migeed’s checks are nominal while ours are structural. We refrain from a performance comparison since Reticulated Python is an interpreter without just-in-time compilation and thus performance tradeoffs are different.

Richards et al. [36] take a similar implementation approach to our work, demonstrating that key mechanisms such as object shapes used by a VM to optimize dynamic languages can be used to eliminate most of the overhead of dynamic type checks. Unlike our work, Richards implement “full” gradual typing with blame tracking, rather than shallow structural checks, and do so on top of an adapted Higgs VM. The Higgs VM implements a baseline just-in-time compiler based on basic-block versioning [20]. In contrast, our implementation of dynamic checks is built on top of the Truffle framework for the Graal VM, and reaches performance approaching that of V8 (cf. section 4.2). The performance difference is of relevance here since any small constant factors introduces into a VM with a lower baseline performance can remain hidden, while they stand out more prominently on a faster baseline.

Overall, it is unclear whether our results confirm the ones reported by Richards et al. [36], because our system is simpler. It does not introduce the polymorphism issues caused by accumulating cast information on object shapes, which could be important for performance. Considering that Richards et al. report ca. 4% overhead on the classic Richards benchmark, while we see 31%, further work seems necessary to understand the performance implications of their approach for a highly optimizing just-in-time compiler.

7 Conclusion

With the wide-spread use of dynamically, optionally, and gradually typed languages, efficient techniques for type checking become more important. In this paper, we have demonstrated that optimizing virtual machines enable dynamic checks of shallow structural types with relatively little overhead, and require only small modifications to an AST interpreter. We evaluated this approach with Moth, an implementation of the Grace language on top of Truffle and Graal.

In our implementation, types are structural and shallow: a type specifies only the names of members provided by objects, and not the types of their arguments and results. These types are checked on access to variables, when assigning to method parameters, and also on return values. The information on types is encoded as part of an object's shape, which means that shape checks already performed in an optimizing dynamic language implementation can be used to check types, too. Being able to tie checks to the shapes in this way is critical for reducing the overhead of dynamic checking.

Using the Are We Fast Yet benchmarks as well as a collection of benchmarks from the gradual typing literature, we find that our approach to dynamic type checking introduces an overhead of 5% (min. -10%, max. 86%) on peak performance. Since Moth reaches the performance of a highly optimized JavaScript VM such as V8, we believe that these results are a good indication for the low overhead of our approach.

In specific cases, the overhead is still significant and requires further research to be practical. Thus, future research should investigate how the number of dynamic type checks can be reduced without causing the type feedback to become too imprecise to be useful. One approach might increase the necessary changes to a language implementation, but avoid checking every variable read. Another approach might further leverage Truffle's self-specialization to propagate type requirements and avoid unnecessary checks.

Finally, we hope to apply our approach to full structural types with blame that support the gradual guarantee. For Vitousek et al. [48]'s transient semantics, this should primarily require finer distinctions in the subtype matrix; monotonic and guarded semantics will require more work, including extensions to the underlying object model. This should let us verify that Richards et al. [36]'s results generalize to highly optimizing virtual machines, or alternatively, show that other optimizations for precise blame need to be investigated.

References

- [1] Martín Abadi, Luca Cardelli, Benjamin C. Pierce, and Gordon D. Plotkin. 1991. Dynamic Typing in a Statically Typed Language. *ACM Trans. Program. Lang. Syst.* 13, 2 (1991), 237–268.
- [2] Amal Ahmed, Robert Bruce Findler, Jeremy G. Siek, and Philip Wadler. 2011. Blame for all. In *Proceedings of the 38th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2011, Austin, TX, USA, January 26–28, 2011*. 201–214.
- [3] Edd Barrett, Carl Friedrich Bolz-Tereick, Rebecca Killick, Sarah Mount, and Laurence Tratt. 2017. Virtual Machine Warmup Blows Hot and Cold. *Proc. ACM Program. Lang.* 1, OOPSLA, Article 52 (Oct. 2017), 27 pages.
- [4] Spenser Bauman, Carl Friedrich Bolz, Robert Hirschfeld, Vasily Kirilichev, Tobias Pape, Jeremy G. Siek, and Sam Tobin-Hochstadt. 2015. Pycket: a tracing JIT for a functional language. In *Proceedings of the 20th ACM SIGPLAN International Conference on Functional Programming, ICFP 2015, Vancouver, BC, Canada, September 1–3, 2015*. 22–34.
- [5] Spenser Bauman, Carl Friedrich Bolz-Tereick, Jeremy Siek, and Sam Tobin-Hochstadt. 2017. Sound Gradual Typing: Only Mostly Dead. *Proc. ACM Program. Lang.* 1, OOPSLA, Article 54 (Oct. 2017), 24 pages.
- [6] Gavin M. Bierman, Martín Abadi, and Mads Torgersen. 2014. Understanding TypeScript. In *ECOOP 2014 - Object-Oriented Programming - 28th European Conference, Uppsala, Sweden, July 28 - August 1, 2014. Proceedings*. 257–281.
- [7] Andrew P. Black, Kim B. Bruce, Michael Homer, and James Noble. 2012. Grace: the absence of (inessential) difficulty. In *Onward! '12: Proceedings 12th SIGPLAN Symp. on New Ideas in Programming and Reflections on Software*. ACM, New York, NY, 85–98. <http://doi.acm.org/10.1145/2384592.2384601>
- [8] Andrew P. Black, Kim B. Bruce, and James Noble. 2010. Panel: designing the next educational programming language. In *SPLASH/OOPSLA Companion*.
- [9] Andrew P. Black, Norman C. Hutchinson, Eric Jul, and Henry M. Levy. 2007. The development of the Emerald programming language. In *Proceedings of the Third ACM SIGPLAN History of Programming Languages Conference (HOPL-III), San Diego, California, USA, 9–10 June 2007*. 1–51.
- [10] Carl Friedrich Bolz, Antonio Cuni, Maciej Fijałkowski, and Armin Rigo. 2009. Tracing the Meta-level: PyPy's Tracing JIT Compiler. In *Proceedings of the 4th Workshop on the Implementation, Compilation, Optimization of Object-Oriented Languages and Programming Systems (ICOOOLPS '09)*. ACM, 18–25.
- [11] Carl Friedrich Bolz, Lukas Diekmann, and Laurence Tratt. 2013. Storage Strategies for Collections in Dynamically Typed Languages. In *Proceedings of the 2013 ACM SIGPLAN International Conference on Object Oriented Programming Systems Languages & Applications (OOPSLA'13)*. ACM, 167–182.
- [12] Carl Friedrich Bolz and Laurence Tratt. 2013. The Impact of Meta-Tracing on VM Design and Implementation. *Science of Computer Programming* (2013).
- [13] John Tang Boyland. 2014. The Problem of Structural Type Tests in a Gradual-Typed Language. In *FOOL*.
- [14] Gilad Bracha. 2004. Pluggable Type Systems. *OOPSLA Workshop on Revival of Dynamic Languages*. , 6 pages.
- [15] Gilad Bracha. 2015. *The Dart Programming Language*. Addison-Wesley Professional.
- [16] Gilad Bracha and David Griswold. 1993. Stongtalk: Typechecking Smalltalk in a Production Environment. In *OOPSLA*.
- [17] Gilad Bracha, Peter von der Ahé, Vassili Bykov, Yaron Kashai, William Maddox, and Eliot Miranda. 2010. Modules as Objects in Newspeak. In *European Conference on Object-Oriented Programming (ECOOP)*. Lecture Notes in Computer Science, Vol. 6183. 405–428.
- [18] Kim Bruce, Andrew Black, Michael Homer, James Noble, Amy Ruskin, and Richard Yannow. 2013. Seeking Grace: a new object-oriented language for novices. In *Proceedings 44th SIGCSE Technical Symposium on Computer Science Education*. ACM, 129–134.
- [19] Craig Chambers, David Ungar, and Elgin Lee. 1989. An Efficient Implementation of SELF a Dynamically-Typed Object-Oriented Language Based on Prototypes. In *Proceedings on Object-Oriented Programming Systems, Languages and Applications (OOPSLA'89)*. ACM, 49–70.

- [20] Maxime Chevalier-Boisvert and Marc Feeley. 2016. Interprocedural Type Specialization of JavaScript Programs Without Type Analysis. In *30th European Conference on Object-Oriented Programming (ECOOP 2016) (LIPIcs)*, Vol. 56. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 7:1–7:24. <https://doi.org/10.4230/LIPIcs.ECOOP.2016.7>
- [21] Daniel Clifford, Hannes Payer, Michael Stanton, and Ben L. Titzer. 2015. Memento Mori: Dynamic Allocation-site-based Optimizations. In *Proceedings of the 2015 International Symposium on Memory Management (ISMM '15)*. ACM, 105–117.
- [22] Benoit Daloze, Stefan Marr, Daniele Bonetta, and Hanspeter Mössenböck. 2016. Efficient and Thread-Safe Objects for Dynamically-Typed Languages. In *Proceedings of the 2016 ACM International Conference on Object Oriented Programming Systems Languages & Applications (OOPSLA'16)*. ACM, 642–659.
- [23] Ulan Degenbaev, Jochen Eisinger, Manfred Ernst, Ross McIlroy, and Hannes Payer. 2016. Idle Time Garbage Collection Scheduling. In *Proceedings of the 37th ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI'16)*. ACM, 570–583.
- [24] M. Furr, J.-H. An, J. Foster, and M.J. Hicks. 2009. Static type inference for Ruby. In *Symposium on Applied Computation*. 1859–1866.
- [25] Adele Goldberg and David Robson. 1983. *Smalltalk-80: The Language and its Implementation*. Addison-Wesley.
- [26] Ben Greenman and Zeina Migeed. 2018. On the Cost of Type-tag Soundness. In *Proceedings of the ACM SIGPLAN Workshop on Partial Evaluation and Program Manipulation (PEPM'18)*. ACM, 30–39.
- [27] Christian Humer, Christian Wimmer, Christian Wirth, Andreas Wöß, and Thomas Würthinger. 2014. A Domain-Specific Language for Building Self-Optimizing AST Interpreters. In *Proceedings of the 13th International Conference on Generative Programming: Concepts and Experiences (GPCE '14)*. ACM, 123–132. <https://doi.org/10.1145/2658761.2658776>
- [28] Ralph E. Johnson. 1986. Type-Checking Smalltalk. In *Conference on Object-Oriented Programming Systems, Languages, and Applications (OOPSLA'86)*, Portland, Oregon, USA, Proceedings. 315–321.
- [29] Timothy Jones, Michael Homer, James Noble, and Kim Bruce. 2016. Object Inheritance Without Classes. In *30th European Conference on Object-Oriented Programming (ECOOP 2016)*, Vol. 56. 13:1–13:26.
- [30] Andre Kuhlenschmidt, Deyaaeldeen Almahallawi, and Jeremy G. Siek. 2018. Efficient Gradual Typing. *CoRR* abs/1802.06375 (2018). arXiv:1802.06375
- [31] Ole Lehrmann Madsen, Birger Möller-Pedersen, and Kristen Nygaard. 1993. *Object-Oriented Programming in the BETA Programming Language*. Addison-Wesley.
- [32] Stefan Marr. 2018. SOMns: A Newspeak for Concurrency Research. <https://github.com/smarr/SOMns>.
- [33] Stefan Marr, Benoit Daloze, and Hanspeter Mössenböck. 2016. Cross-Language Compiler Benchmarking—Are We Fast Yet?. In *Proceedings of the 12th Symposium on Dynamic Languages (DLS'16)*. ACM, 120–131.
- [34] Stefan Marr and Stéphane Ducasse. 2015. Tracing vs. Partial Evaluation: Comparing Meta-Compilation Approaches for Self-Optimizing Interpreters. In *Proceedings of the 2015 ACM International Conference on Object Oriented Programming Systems Languages & Applications (OOPSLA '15)*. ACM, 821–839.
- [35] Fabian Muehlboeck and Ross Tate. 2017. Sound Gradual Typing is Nominally Alive and Well. *Proc. ACM Program. Lang.* 1, OOPSLA, Article 56 (Oct. 2017), 30 pages.
- [36] Gregor Richards, Ellen Artca, and Alexi Turcotte. 2017. The VM Already Knew That: Leveraging Compile-time Knowledge to Optimize Gradual Typing. *Proc. ACM Program. Lang.* 1, OOPSLA, Article 55 (Oct. 2017), 27 pages.
- [37] Richard Roberts, Stefan Marr, Michael Homer, and James Noble. 2017. Toward Virtual Machine Adaption Rather than Reimplementation. In *MoreVMs'17: 1st International Workshop on Workshop on Modern Language Runtimes, Ecosystems, and VMs at <Programming> 2017*. Presentation.
- [38] Jeremy G Siek and Walid Taha. 2006. Gradual typing for functional languages. In *Scheme and Functional Programming Workshop*.
- [39] Jeremy G. Siek, Michael M. Vitousek, Matteo Cimini, and John Tang Boyland. 2015. Refined Criteria for Gradual Typing. In *1st Summit on Advances in Programming Languages, SNAPL 2015, May 3-6, 2015, Asilomar, California, USA*. 274–293.
- [40] Jeremy G. Siek, Michael M. Vitousek, Matteo Cimini, Sam Tobin-Hochstadt, and Ronald Garcia. 2015. Monotonic References for Efficient Gradual Typing. In *European Symposium on Programming (ESOP)*. 432–456.
- [41] Jeremy G. Siek and Philip Wadler. 2010. Threesomes, with and without blame. In *Proceedings of the 37th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2010, Madrid, Spain, January 17-23, 2010*. 365–376.
- [42] Vincent St-Amour, Sam Tobin-Hochstadt, and Matthias Felleisen. 2012. Optimization coaching: optimizers learn to communicate with programmers. In *Proceedings of the 27th Annual ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications, OOPSLA 2012, part of SPLASH 2012, Tucson, AZ, USA, October 21-25, 2012*. 163–178.
- [43] G.L. Steele. 1990. *Common Lisp the Language*. Digital Press.
- [44] Nataliia Stulova, José F. Morales, and Manuel V. Hermenegildo. 2016. Reducing the Overhead of Assertion Run-time Checks via Static Analysis. In *Proceedings of the 18th International Symposium on Principles and Practice of Declarative Programming (PPDP'16)*. ACM, 90–103.
- [45] Asumu Takikawa, Daniel Feltey, Ben Greenman, Max S. New, Jan Vitek, and Matthias Felleisen. 2016. Is Sound Gradual Typing Dead?. In *Proceedings of the 43rd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL'16)*. ACM, 456–468.
- [46] The Clean. 1990. Vehicle. Flying Nun Records, FN147.
- [47] Sam Tobin-Hochstadt and Matthias Felleisen. 2008. The design and implementation of Typed Scheme. In *Proceedings of the 35th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2008, San Francisco, California, USA, January 7-12, 2008*. 395–406.
- [48] Michael M. Vitousek, Andrew M. Kent, Jeremy G. Siek, and Jim Baker. 2014. Design and evaluation of gradual typing for Python. In *DLS'14, Proceedings of the 10th ACM Symposium on Dynamic Languages, part of SPLASH 2014, Portland, OR, USA, October 20-24, 2014*. 45–56.
- [49] Michael M. Vitousek, Cameron Swords, and Jeremy G. Siek. 2017. Big Types in Little Runtime: Open-world Soundness and Collaborative Blame for Gradual Type Systems. In *Proceedings of the 44th ACM SIGPLAN Symposium on Principles of Programming Languages (POPL'17)*. ACM, 762–774.
- [50] Philip Wadler and Robert Bruce Findler. 2009. Well-Typed Programs Can't Be Blamed. In *European Symposium on Programming Languages and Systems (ESOP)*. 1–16.
- [51] Andreas Wöß, Christian Wirth, Daniele Bonetta, Chris Seaton, Christian Humer, and Hanspeter Mössenböck. 2014. An Object Storage Model for the Truffle Language Implementation Framework. In *Proceedings of the 2014 International Conference on Principles and Practices of Programming on the Java Platform: Virtual Machines, Languages, and Tools (PPPJ'14)*. ACM, 133–144.
- [52] Thomas Würthinger, Christian Wimmer, Christian Humer, Andreas Wöß, Lukas Stadler, Chris Seaton, Gilles Duboscq, Doug Simon, and Matthias Grimmer. 2017. Practical Partial Evaluation for High-performance Dynamic Language Runtimes. In *Proceedings of the 38th ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI'17)*. ACM, 662–676.
- [53] Thomas Würthinger, Christian Wimmer, Andreas Wöß, Lukas Stadler, Gilles Duboscq, Christian Humer, Gregor Richards, Doug Simon, and Mario Wolczko. 2013. One VM to Rule Them All. In *Proceedings of the 2013 ACM International Symposium on New Ideas, New Paradigms, and Reflections on Programming & Software (Onward! 2013)*. ACM, 187–204.

- [54] Thomas Würthinger, Andreas Wöß, Lukas Stadler, Gilles Duboscq, Doug Simon, and Christian Wimmer. 2012. Self-Optimizing AST Interpreters. In *Proceedings of the 8th Dynamic Languages Symposium (DLS'12)*. 73–82. <https://doi.org/10.1145/2384577.2384587>
- [55] Carlo Zapponi. 2018. GitHub. [GitHub.info](https://github.com). Accessed 29 June 2018.