



NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

SCHOOL OF SCIENCES

DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS

POSTGRADUATE STUDIES PROGRAM

MASTER THESIS

Faster Scala Collections with Macros

Georgios Kollias

Supervisor: Yannis Smaragdakis, Associate Professor NKUA

ATHENS

MAY 2013



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Γρηγορότερες Δομές Δεδομένων στη Scala
με χρήση Macros**

Γεώργιος Κόλλιας

Επιβλέπων: Γιάννης Σμαραγδάκης, Αναπληρωτής Καθηγητής ΕΚΠΑ

ΑΘΗΝΑ

ΜΑΙΟΣ 2013

MASTER THESIS

Faster Scala Collections with Macros

Georgios Kollias

RN: M1049

SUPERVISOR:

Yannis Smaragdakis, Associate Professor NKUA

THESIS COMMITTEE:

Yannis Smaragdakis, Associate Professor NKUA

Panos Rondogiannis, Associate Professor NKUA

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Γρηγορότερες Δομές Δεδομένων στη Scala
με χρήση Macros**

Γεώργιος Κόλλιας

ΑΜ: M1049

ΕΠΙΒΛΕΠΩΝ :

Γιάννης Σμαραγδάκης, Αναπληρωτής Καθηγητής ΕΚΠΑ

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:

Γιάννης Σμαραγδάκης, Αναπληρωτής Καθηγητής ΕΚΠΑ

Παναγιώτης Ροντογιάννης, Αναπληρωτής Καθηγητής ΕΚΠΑ

Περίληψη

ΦιΞμε
Φαταλ:
Ρεπλασε
με

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: ...

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: ...

Abstract

...

SUBJECT AREA: ...

KEYWORDS: ...

Acknowledgements

Fixme
Fatal:
Add me

Contents

1	Introduction	14
2	Background	17
2.1	Scala Collections Overview	17
2.2	Scala Compile-Time Reflection Overview	18
3	Our Approach: ft-declosurify	22
3.1	FT-DECLOSURIFY Overview	22
3.2	FT-DECLOSURIFY Implementation Specifics	25
4	Safe Publication	33
5	Experimental Results	34
6	Related Work	35
7	Conclusions	36
	Acronyms and Abbreviations	37
	Appendices	38
A	Escape Analysis Code	38

List of Figures

1.1	Example of generic map in Java	15
2.1	All scala.collection collections	18
3.1	scalac's internal representation of <code>List(1, 2, 3).map(_ + 1)</code>	22
3.2	Scala's default map implementation	23
3.3	<code>List(1, 2, 3).macroMap(_ + 1)</code> example transformation	24

List of Tables

List of Corrections

Fatal: Replace me	5
Fatal: Add me	7
Fatal: Replace me	13
Fatal: add abbr	14
Fatal: cite	14
Fatal: cite http://www.infoq.com/news/2011/11/yammer-scala	16
Fatal: cite	17
Fatal: cite	17
Fatal: cite	17
Fatal: date	18
Fatal: cite Template Haskell	19
Fatal: cite Nemerle	19
Fatal: abbr	20
Fatal: cite 19, 8	20
Fatal: add the respective slide from the presentation	25
Fatal: Replace me	33
Fatal: Replace me	34
Fatal: Replace me	35
Fatal: Replace me	36
Fatal: Add me	38

Todo list

Preface

Fixme
Fatal:
Replace
me

Chapter 1

Introduction

This chapter will introduce our work and the problem it tries to attack. It is a real-world hard problem that affects the whole Java Virtual Machine (JVM) ecosystem. This work tries to mitigate many of its implications. In the next chapter, we will have an overview of the Scala language, its standard library collections and its new compile-time reflection subsystem in order to prepare us for Chapter X where we will explain our project’s implementation and core functionality. Chapter X provides us with several benchmarks, showing promising performance speedups. Chapter X presents some similar work in the area. Finally, Chapter X summarizes this work.

Section: The Problem and Our Approach

In “Fixing The Inlining Problem”, Cliff Click describes an issue that has emerged the last few years in the JVM ecosystem:

“*The Problem* is simply this: new languages on the JVM (e.g., JRuby) and new programming paradigms (e.g., Fork Join) have exposed a weakness in the current crop of inlining heuristics. Inlining is not happening in a crucial point in hot code exposed by these languages, and the lack of inlining is hurting performance in a major way. AND the inlining isn’t happening because The Problem is a hard one to solve; (i.e. it’s not the case that we’ll wave our wands and do a quick fix & rebuild HotSpot and the problem will go away). John Rose, Doug Lea and I all agree that it’s a Major Problem facing JVMs with long and far reaching implications.”

Let’s see what *The Problem* is with a small example in a Java-like language. The Problem is getting the right amount of context in hot inner loops - which also contain a *megamorphic* virtual call in the loop and not much else. Here is a naive implementation of a `map` method that applies a predetermined function to all the elements of a source array and assigns the result to a destination array. In this case we increment all source’s elements by one:

```

1 // The function in the inner loop
2 long add1(long a) {return a + 1;}
3 // The iterator function
4 void map(long[] dst, long[] src) {
```

fixme
Fatal:
add abbr

fixme
Fatal:
cite

```

1 // A sample iterator function
2 void map( CallableOneArg fcnlrg, long[] dst, long[] src) {
3     for(int i=0; i < dst.len; i++)
4         dst[i] = fcnlrg.call(src[i]);
5 }

```

Figure 1.1: Example of generic map in Java

```

5     for(int i=0; i < dst.len; i++) // simple loop
6         dst[i] = add1(src[i]); // around a simple loop body
7 }

```

Inlining the function `add1` is crucial to performance here. Without inlining the compiler does not know what the loop body does (because function calls can in general do anything), and with inlining it can understand the entire function completely - and then see it's a simple loop around a stream of array references. At this point the JIT can do range-check elimination, loop unrolling, and prefetching, among other optimizations.

The Problem is that there are multiple variations of `add1` *and* the wrapping iterator gets complex. It's the product of these two parts getting complicated that makes The Problem. In this work, we are mostly interested in the first part, since the Scala collections iterators are not very big or complex.

More often than not, after implementing `map`, we would like to add more functions similar to `add1`. We might also want to add `add2`, `mult3`, `filter` and so on. What we really want is a way to pass in the function to apply on the basic data bits in the innermost loop of our iterator.

In Java we often do this with either a `Callable` or a `Runnable`:

We need only 1 copy of our iterator, and we can apply nearly all kinds of one argument functions. Alas, that inner loop now contains a function call that needs inlining and there are dozens of different functions for `fcn1rg.call`. The JIT does not know which one to inline here, because all these different functions are called at different times. Typically then the JIT does not inline any of them, and instead opts for a virtual call. While the virtual call itself is not too bad, the lack of knowledge of what goes on inside the virtual call prevents all kinds of crucial optimizations: loop unrolling, range-check elimination, all kinds of prefetching and alias analyses.

One solution would be to make the inner function call a static (final) call, which then the JIT can inline. Of course, if we do that we need an iterator for the `add1` version, one for the `add2` version, and one for the `mult3` version, so we need a lot of them. Also, we will need a new one for each new function we can think of; we cannot just name them all up front. So we will end up with a lot of these iterators each with a custom call in the inner loop. All these iterators will start blowing out the instruction cache on our CPU, and besides it is a pain to maintain dozens of cloned possibly complex iterators.

Several of the Java ecosystem's prominent figures, like Cliff Click, John Rose, Doug Lea, have proposed solutions that range from pure obscure technical to pure educational ones, demonstrating possible *megamorphic inlining friendly* coding styles.

Scala, as most of the modern JVM languages, is no exception and it suffers from The Problem too and, actually, it affects its adoption negatively. At the end of 2011, an email from a Yammer employee towards the CEO of TypeSafe, the company backing the Scala ecosystem, about Scala shortcomings, leaked to the public . Most complaints were related with The Problem and, more specifically, with the Scala standard library collections' performance.

In the next chapters we will see where exactly the problem lies and how our project can help us alleviate it. We have named our project FT-DECLOSURIFY; the `ft` prefix stands for Scala compiler's FastTrack mechanism, as we will see in Chapter X, and the `declosurify` suffix suggests that the implementation is based on Paul Phillips's original DECLOSURIFY project `fxfatalcite` github url.

`ft-declosurify` defines two methods, `macroMap` and `macroForeach`, that can be used from all the Scala standard library collections. They offer the same functionality as their Scala standard library counterparts, `map` and `foreach` methods, but they are much faster because they suffer much less from The Problem. In most cases, `macroMap`/`macroForeach` can be considered as faster and drop-in replacements of the default `map`/`foreach` methods.

Our solution tries to attack The Problem using Scala's new compile-time reflection subsystem. The next chapter will give us an overview of both the Scala collections and the new compile-time reflection capabilities.

Fixme
Fatal:
cite
<http://www.scala>

Chapter 2

Background

Scala is a relatively new statically typed programming language that tries to unify the object-oriented and functional programming paradigms into one coherent paradigm, recently called object-functional. Currently, its main implementation runs on the JVM and so its main goal is to provide a more general and uniform superset of Java. Since version 2.8, Scala has a rich collections library and since version 2.10 it has a completely new reflection subsystem.

Fixme
Fatal:
cite

Fixme
Fatal:
cite

Fixme
Fatal:

cite

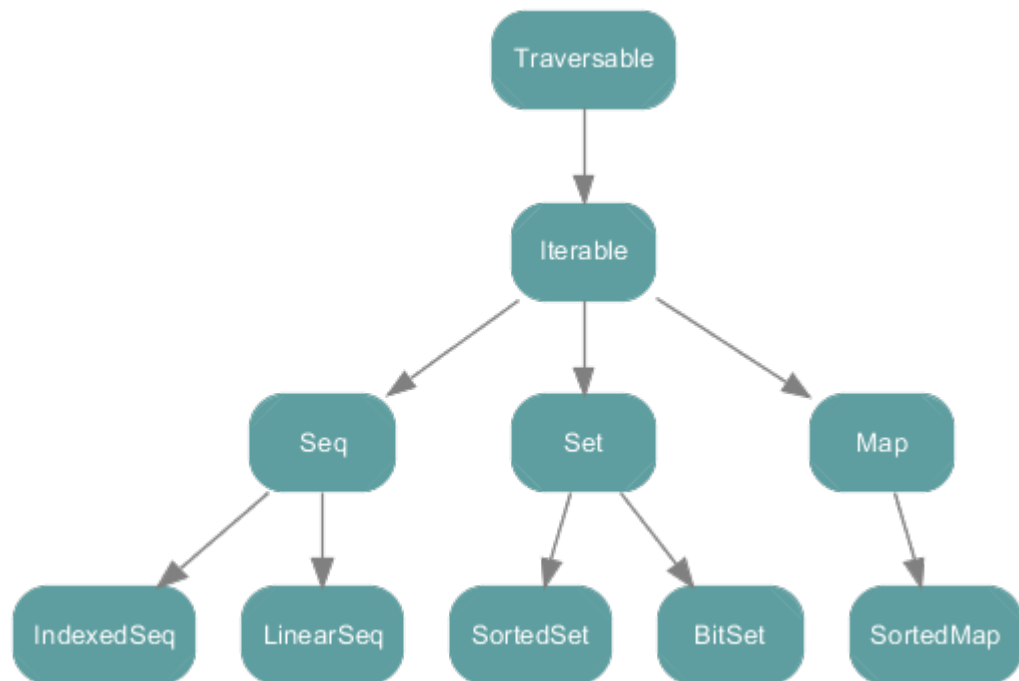
2.1 Scala Collections Overview

The Scala library systematically distinguishes between mutable and immutable collections. A mutable collection can be updated or extended in place. This means one can change, add, or remove elements of a collection as a side effect. Immutable collections, by contrast, never change. We still have operations that simulate additions, removals, or updates, but these operations will in each case return a new collection and leave the old collection unchanged. We will see in the next chapter how this mutable-immutable separation affects our macro transformation plan.

All collection classes are found in the package `scala.collection` or one of its sub-packages `mutable`, `immutable`, and `generic`. Most collection classes needed by client code exist in three variants, which are located in packages `scala.collection`, `scala.collection.immutable`, and `scala.collection.mutable`, respectively. Each variant has different characteristics with respect to mutability.

A collection in package `scala.collection.immutable` is guaranteed to be immutable for everyone. Such a collection will never change after it is created. A collection in package `scala.collection.mutable` is known to have some operations that change the collection in place.

A collection in package `scala.collection` can be either mutable or immutable. For instance, `collection.IndexedSeq[T]` is a superclass of both `collection.immutable.IndexedSeq[T]` and `collection.mutable.IndexedSeq[T]`. Generally, the root collections in package `scala.collection` define the same interface as the immutable collections, and the mutable collections in pack-

Figure 2.1: Basic collections in `scala.collection`

age `scala.collection.mutable` typically add some side-effecting modification operations to this immutable interface.

By default, Scala always picks immutable collections. For instance, if we just write `Set` without any prefix or without having imported `Set` from somewhere, we get an immutable set because these are the default bindings imported from the `scala` package. To get the mutable default version, we need to write explicitly `collection.mutable.Set`.

Figure 2.1 shows all collections in package `scala.collection`. These are all high-level abstract classes or traits, which generally have mutable as well as immutable implementations.¹

This work focuses mainly on `Seq`'s subtree, since it's where we can get the most prominent speedups by exploiting the sequences' properties. A sequence is a kind of iterable that has a `length` method and whose elements have fixed index positions, starting from 0.

2.2 Scala Compile-Time Reflection Overview

Scala version 2.10, released on , introduced a new reflection subsystem adding both run

¹Figure courtesy of Matthias Doenitz

time and compile metaprogramming capabilities. The new run-time reflection is much more general and feature complete compared to Java's reflection. Compile-time reflection is quite rare in mainstream statically typed programming languages and, currently, it can only be found in more exotic languages like Haskell and Nemerle . Compile-time reflection enabled the introduction of an experimental version of type-safe syntactic macros cite Scala Macros, a Technical Report.

Syntactic macro systems work at the level of abstract syntax trees and preserve the lexical structure of the original program. Macro systems that work at the level of lexical tokens, like the C preprocessor, cannot preserve the lexical structure reliably. The most widely used implementations of syntactic macro systems are found in Lisp-like languages such as Common Lisp, Scheme. These languages are especially suited for this style of macro due to their uniform, parenthesized syntax (known as S-expressions).

Compile-time metaprogramming is a valuable tool for enabling such programming techniques as:

- Language virtualization (overloading/overriding semantics of the original programming language to enable deep embedding of DSLs)
- Program reification (providing programs with means to inspect their own code)
- Self-optimization (self-application of domain-specific optimizations based on program reification)
- Algorithmic program construction (generation of code that is tedious to write with the abstractions supported by a programming language)

This work falls in categories three and four, since we use compile-time reflection to generate code programmatically for each `macroMap`/`macroForeach` method, specialized at the call-site for optimization reasons.

Scala's compile-time metaprogramming can be used through Scala's new macro system that allows programmers to write *macro defs*: functions that are transparently loaded by the compiler and executed during compilation.

Our project is implemented directly in the Scala compiler (scalac), so we can use most of the available compile-time metaprogramming capabilities directly, without using macro defs explicitly. In the next chapter we will see how we achieve it.

Fixme
Fatal:
cite
Template
Haskell
Fixme
Fatal:
cite
Nemerle

Compile-time reflection allows us to create new and/or manipulate existing abstract-syntax trees (ASTs) during the compiler's typechecking phase. All the new or changed ASTs are re-typechecked, guaranteeing us type-safe transformations. `scalac` represents ASTs with objects of type `scala.reflect.api.Tree` or `scala.reflect.api.Exprs`, which is just a typed-wrapper of `scala.reflect.api.Tree`. Through the available reflection APIs, we can create, inspect or change the compiler's `scala.reflect.api.Symbols` and `scala.reflect.api.Types` objects that are related with these ASTs.

FixMe
Fatal:
abbr

For example, we can create the AST of the Scala expression `x < 10` manually, either with a macro or directly within `scalac`, with this code `Apply(Select(Ident(newTermName("x")), newTermName("$less"), List(Literal(Constant(10)))))`. `Apply`, `Select`, `Ident`, `Literal`, `Constant` are AST objects themselves of `scala.reflect.api.Tree` type.

Obviously the AST construction is cumbersome and error-prone. But most probably it is also wrong. If the AST was generated within an internal `scalac` method or within a macro, the returned AST will be inlined and type-checked at the method/macro call site. But this means that the identifier `x` will be type-checked at a point where it is most likely not visible, or in the worst case they might refer to something else. In the macro literature, this insensitivity to bindings is called non-hygienic. Scala's compile-time reflection solves the non-hygiene problem providing a built-in macro, called `reify`, that produces its tree one stage later.

FixMe
Fatal:
cite 19, 8

The `reify` macro plays a crucial role in the compile-time metaprogramming. Its definition as a member of `Context` is:

```
1 def reify[T](expr : T): Expr[T] = macro . . .
```

`Reify` accepts a single parameter `expr`, which can be any well-typed Scala expression, and creates a tree that, when compiled and evaluated, will recreate the original tree `expr`. So `reify` is like time-travel: trees get re-constituted at a later stage. If `reify` is called from normal compiled code, its effect is that the AST passed to it will be recreated at run time. Consequently, if `reify` is called from a macro implementation or a method inside `scalac`, its effect is that the AST passed to it will be recreated at macro-expansion time (which corresponds to run time for macros). This gives a convenient way to create syntax trees from Scala code: pass the Scala code to `reify`, and the result will be a syntax tree that represents that very same code.

For example, `reify(x < 10)` will generate an `Expr` object representing the same AST we created manually before.

More importantly, `reify` packages the result expression tree with the types and values

of all free references that occur in it. This means in effect that all free references in the result are already resolved, so that re-typechecking the tree is insensitive to its environment. All identifiers referred to from an expression passed to `reify` are bound at the definition site, and not re-bound at the call site. As a consequence, macros that generate trees only by means of passing expressions to `reify` are hygienic.

So, in a sense, Scala macros are self-cleaning. Their basic form is minimal and unhygienic, but that simple form is expressive enough to formulate a `reify` macro, which in turn can be used to make tree construction in macros concise and hygienic.

Another important compile-time metaprogramming operation is the *splicing*, which could be described as `reify`'s inverse operation. Using `Expr`'s `splice` method we can inject an existing AST inside a `reify`'s body.

Reification and splicing operations are crucial to our implementation, as we will see in the next chapter.

Chapter 3

Our Approach: FT-DECLOSURIFY

This chapter presents the core of our work. Section 3.1 will give us a thorough overview of FT-DECLOSURIFY and section 3.2 will describe the major FT-DECLOSURIFY implementation details.

3.1 FT-DECLOSURIFY Overview

For understanding at a high-level what FT-DECLOSURIFY does let's see how an example expression `List(1, 2, 3).map(_ + 1)` is translated from both the scalac and the FT-DECLOSURIFY points of view. `List(1, 2, 3).map(_ + 1)` is a shortcut for `List(1, 2, 3).map(x => x + 1)`, where `x => x + 1` is an anonymous function -closure- that returns its argument incremented by one. Applying that function to the `List(1, 2, 3)` will result in a new list `List(2, 3, 4)`. Since Scala code is translated into Java bytecode at the end and since Java doesn't support any notion of functions inherently, this anonymous function should be translated somehow in constructs that are supported by the Java bytecode. This trivial Scala expression `List(1, 2, 3).map(_ + 1)` is translated internally by scalac to: ¹

We can see that scalac converts the `_ + 1` function into a block of code (piece of code between two braces), where a class, called `$anonfun`, is defined. That class extends the `AbstractFunction1[Int,Int]` class, which is an abstract class that represents the func-

¹The code listing below as well as most of the following listings are slightly abbreviated for readability reasons

```

1 immutable.this.List.apply[Int](Array[Int]{1, 2, 3}).map[Int, List[Int]]({
2     final class $anonfun extends scala.runtime.AbstractFunction1[Int,Int] with
        Serializable {
3         def <init>(): anonymous class $anonfun = {
4             $anonfun.super.<init>();
5             ()
6         };
7         final def apply(x$1: Int): Int = x$1.+(1)
8     };
9     (new $anonfun(): Int => Int)
10 }, immutable.this.List.canBuildFrom[Int]())

```

Figure 3.1: scalac's internal representation of `List(1, 2, 3).map(_ + 1)`

```

1 def map[B, That](f: A => B)(implicit bf: CanBuildFrom[Repr, B, That]): That =
2 {
3     def builder = {
4         val b = bf(repr)
5         b.sizeHint(this)
6         b
7     }
8     val b = builder
9     for (x <- this) b += f(x)
10    b.result
11 }

```

Figure 3.2: Scala’s default map implementation

tions that take one integer argument and return another integer. Inside the class an `apply` method is defined which is called whenever we apply a class object to one integer argument. The `apply` body returns its argument incremented by one, e.g., `val a1 = new $anonfun(); a1(5);` returns 6. Just after the class definition, scalac creates a new object of this class and this is what is actually returned from that block of code. Eventually, the `_ + 1` is substituted by an object of a subclass of a class representing functions internally, which leads us to the conclusion that the scala source-level closures are translated to function class objects. Below we attach the standard Scala library’s basic map implementation:

Here `f` is the function object scalac passed during the `map` call, i.e., the object resulted from `new $anonfun()` call in Figure 3.1. `f(x)` invocation will be expanded to `f.apply(x)` and, therefore, it is a normal method call on object `f`. We can easily see how similar this map definition is with the one provided in Figure 1.1. They both suffer from The Problem. Here, `f`’s runtime type will, usually, be different on each `map` call since the functions we pass are generally different. Such calls are called mega-morphic virtual calls and, currently, cannot be inlined by the JVM as we explained in the introduction. So on each `map` we generally have the added overhead of a dynamic call to the passed function object. Even worse, the lack of knowledge of what goes on inside the virtual call prevents all kinds of crucial optimizations.

Let’s see how FT-DECLOSURIFY translates the respective piece of code `List(1, 2, 3).macroMap(_ + 1)`. The only difference here is the use of `macroMap` instead of the plain `map`. It’s transformed into:

Here we use a `scala.collection.mutable.Builder` object to construct the target collection. The builder itself is created from the `canBuildFrom` object the compiler passed in implicitly, as we will explain more thoroughly in section 3.2. The builder creation and initialization takes

```

1 {
2   private def local1(x\$: Int): Int = x\$.+(1);
3   private def builder1: scala.collection.mutable.Builder[Int, List[Int]] = {
4     val b: scala.collection.mutable.Builder[Int, List[Int]] =
5     immutable.this.List.canBuildFrom[Int].apply();
6     b.sizeHint(immutable.this.List.apply[Int](1, 2, 3));
7     b
8   };
9   val buf = builder1();
10  var these = immutable.this.List.apply[Int](1, 2, 3);
11  while (!these.isEmpty) {
12    buf.+=(local1(these.head));
13    these = these.tail
14  };
15  buf.result
16 }

```

Figure 3.3: `List(1, 2, 3).macroMap(_ + 1)` example transformation

place in the `builder1` method. While we traverse the prefix, we apply the `local1` function to all elements and we append (through method `+=`) them to the builder. When the traversal is over, we call the builder's `result` method which returns the full target collection we want (a `List` in this example).

In `macroMap`'s transformation, we see that there is no explicit call to any `map` method. Instead, the list's traversal happens directly within a `while` loop where the `local1` local free method is applied to all of the list's elements. The `local1` method is called a free method since it's not directly attached to any class or object. During `scalac`'s "flatten" phase, it will be lifted and become a method of the enclosing class with a new mangled name. That way `local1`'s receiver runtime type will, likely, remain unchanged during a program's execution. Therefore, it's much easier for JVM to inline `local1` and reason about further optimizations.

We can easily see that `FT-DECLOSURIFY` works well for cases where the closure is statically available at the call-site, like `List(1, 2, 3).macroMap(x => x + 1)`, where it will be expanded to something like

```

1 {
2   def local1(x\$: Int): Int = x\$.+(1);
3   ...
4 }

```

according to the transformations above. This case doesn't solve The Problem at its core but, instead, it "sidesteps" it at a "metalinguistic" level, by eager compile-time inlining and by

transforming the passed function object to a local method.

A more interesting case is when we have something like `List(1, 2, 3).macroMap(myf)`, where we don't know much about the `myf` function except, let's say, its type is `Int => Int`. Then, `macroMap` will be expanded to something like:

```

1 {
2   def local1(x:$1: Int): Int = myf.apply(x:$1);
3   ...
4 }
```

Initially, it seems that we fall back to The Problem without having any advantage, since `myf.apply` seems to be a megamorphic virtual call again. But, in a program, all `macroMap` instances will be expanded at the call-site and their `myf`'s runtime type will, usually, remain the same during each call and, theoretically, the type profiler can inline each `macroMap`'s `myf.apply`, which again alleviates The Problem. So in both cases, we do achieve some wins against The Problem.

In summary, `macroMap` transformation takes advantage of:

- Knowing the static type of `macroMap`'s receiver, because it can apply different transformations depending on the type.
- Knowing the static type of the `macroMap`'s function, because it can transform it in a local method and, so, it's easier for JVM to inline it.
- Having a fixed built-in implementation, applying eager compile-time inlining.

3.2 FT-DECLOSURIFY Implementation Specifics

The `macroForeach` method, just like the standard `foreach`, is meant to traverse all elements of the collection, and apply the given operation, `f`, to each element. The invocation of `f` is done for its side effect only; in fact any function result of `f` is discarded. Similarly, `macroMap`, just like the standard `map`, traverses all elements of the collection, and applies the given operation, `f`, to each element but, also, it always returns a collection whose type depends on `f`. We could see `macroMap`'s functionality as a superset of `macroForeach`'s functionality and that's true on the implementation side too. The same code handles both cases but, in the case of `macroForeach`, we ignore the result. For that reason and without any loss of generality we will be referring only to `macroMap` in the rest of the text.

`macroMap` is defined in `scala.collection.TraversableLike` trait, where the standard library's `map` method is defined too. Its definition is an unusual one:

```
“def macroMap[B, That](f0: A => B)(implicit bf: CanBuildFrom[Repr, B, That]): That =
???”
```

Instead of a usual implementation on the right hand side, we see a `???`. The `???` is an actual special Scala method which is used to throw `NotImplementedError` exceptions. Our `macroMap` implementation will be generated at compile-time, as we will explain in the next subsections, so no exception is raised. It's here just to prevent syntax errors.

The above signature is exactly the same as the `map`'s one. Regarding type parameters, `A` is the collection's element type and, therefore, `f0`'s argument type, `B` is `f0`'s return type, `Repr` is the underlying collection's type and `That` is the generated collection's type. Also, as we said, `macroMap` is a curried method. The first parameter list accepts the operation's function. The second parameter gets an implicit `CanBuildFrom` object. `CanBuildFrom` objects are builder factories which generate the appropriate `scala.collection.generic.Builder` objects depending on their type parameters. Generally, builders makes the creation of a new collection out of existing ones easier and more maintainable. For example, a `CanBuildFrom[List[String], Int, Array[Int]]` object will create a `Builder[Int, Array[Int]]` object which can create an `Array` of integers out of a `List` of strings. Since `bf` is implicit, it will usually be generated and passed by the compiler automatically.

Subsection: FT-DECLOSURIFY in the Scala standard library

The upstream declosurify project can be used as a library adding the -extension- methods `macroMap` and `macroForeach` on all Scala collections that implement the `scala.collection.Traversable` trait and `Arrays`, through Scala's implicit conversions cite implicit conversions. Since, ultimately, we would like to substitute the default `map` and `foreach` methods with the `macroMap` and `macroForeach` implementations, respectively, we had to move the definitions of `macroMap`, `macroForeach` inside the Scala standard library. The Scala standard library doesn't have access to the new compile-time reflection capabilities since it doesn't depend on the `scala-reflect` and `scala-compiler` packages, so we cannot have macro definitions in it. As a solution, we used the scala compiler's `FastTrack` mechanism defined in the `scala.tools.reflect.FastTrack` trait.

`FastTrack` is a low level mechanism of `scalac` to invoke macro expansions for builtin macros and it is, also, the key machinery used by our implementation in order to register and invoke our transformations. One builtin macro that uses this mechanism is the `reify` function itself that we saw in the previous chapter. The `FastTrack` module uses a registry

where one links a method's compiler Symbols with special handler methods. For example, whenever the compiler sees a Symbol representing an application of `reify`, it calls its respective FastTrack special handler which has access to the `reify` application's call-site context and all of its arguments' Abstract Syntax Trees (ASTs), through pattern matching. The call-site context object, generally, holds call-site information like the macro application's enclosing method, its enclosing class, its line in the file. The ASTs contain the internal representation of the macro application like `List(1, 2, 3).macroMap(_ + 1)` i.e., the receiver object, the `macroMap` method call and its function argument. Pattern matching happens upon these ASTs so, theoretically, we can choose to match successfully against `List(1,2,3).macroMap(_ + 1)` but not against `val r = List(1,2,3); r.macroMap(_ + 1)` depending on our needs. If the pattern matching is successful, its respective compiler handler will eventually generate the AST which replaces the macro application's AST at the call-site, completing that way the low-level macro expansion. In our case, whenever `scalac` finds a Symbol of an application of the `scala.collection.TraversableLike` trait's `macroMap` method, it triggers the expansion we will describe in the next subsection.

Regular Scala macros implementations can be generic in the sense that they can be customized with type parameters as any regular Scala method. On top of that, the Scala compiler allows us to “tag” each of these type parameters with special compiler-generated objects, of type `(Weak)TypeTag` [cite], that store the type parameters' full types on each call and make them available at runtime. In short, it's Scala's solution against JVM's type erasure. This machinery is especially useful for macros, since we can inspect these `TypeTag` objects and generate the most suitable code for each occasion. Unfortunately, using the low level compiler's FastTrack mechanism doesn't permit using this facility. Instead, we are forced to work directly with the compiler's internal type representations lowering the level of abstraction we can work with.

FastTrack's special handler for the `macroMap` method, after doing some preprocessing on the pattern matched ASTs, will call the method that does the actual transformation we saw in the previous subsection.

Subsection: Transformation Method Interface

Firstly, let's see the transformation method's signature and what arguments it obtains from the handler. This method is defined in `scala.tools.reflect.declosurify.Declosurify` object `“def mapInfix[A, B](c0: Ctx)(f0: c0.Expr[A => B], inElemTpe: c0.Type, outElemTpe: c0.Type, inCollTpe: c0.Type, outCollTpe: c0.Type, bfTree: c0.Tree): c0.Tree”`. Despite its name, `mapInfix` can generate ASTs for both `macroMap` and `macroForeach`. It's a curried method

parameterized on the received function's input (A) and return (B) types. In the first parameter list, `c0` is a context object as we described it above, giving us access to the macro application's call-site information. The most important field of `c0` is the "prefix" which represents the receiver collection object of the `macroMap` application. We will call it simply "prefix" from now on. In the second parameter list, `f0` represents the function's AST which is going to be applied to the prefix. Its type "Expr" wraps an AST and an internal type tag ("TypeTag") to provide access to the type of the tree. As we mentioned before, using "Expr" doesn't really help us working at that level of the compiler, since we cannot really exploit the (Weak)TypeTag's facilities. The next four arguments are: - `inElemTpe`: the internal compiler type of the prefix's elements - `outElemTpe`: the internal compiler type of the elements of the collection our `macroMap` is going to return - `inCollTpe`: the internal compiler type of the prefix - `outCollTpe`: the internal compiler type of the collection our `macroMap` is going to return

The last element is the AST of the implicit `scala.collection.generic.CanBuildFrom` [cite adrian moors phd] object that got inserted by the compiler at the `macroMap` call-site. This provides us with an easy and accurate way to create the appropriate builder for the generated collection taking advantage of all the existing implicit `CanBuildFrom` objects that are declared in the standard library.

It's easy to observe that all of second parameter list types are prepended with the context object `c0`. This qualified notation realizes the notion of Scala's path-depended types. For example, here we choose the "Expr" that is defined inside the `c0` object. Generally, if we have two different objects `c0` and `c1` of the same type which include an inner type `MyType`, e.g., through type member or inner class, then `c0.MyType` is a different type from `c1.MyType` in Scala. The same holds for the "`c0.Type`"s that follow.

Subsection: Transformation Requirements

Knowing what our transformation method can operate upon we can examine the main points of the actual transformation. One of the first things FT-DECLOSURIFY checks is if the "outCollTpe" is Unit or not (lines XX-YY). If it's Unit then it means the macro is applied only for its side-effects so it's a `foreach` call. Instead of relying on this heuristic, newer version of the Scala macros provide us with the exact name of the calling method.

After that, we check three conditions to ensure that a typical transformation can take place (lines XX-YY): - the `f0` AST is actually a function AST or a block whose last expression is a function, since, in Scala, the value of a block is the block's last expression. - the `f0` AST doesn't contain any return expressions. - macro application is enclosed in a method.

Currently, that limitation makes the transformation easier.

If any of the above points is not satisfied, then the `mapInfix` falls back to the default `map/foreach` implementation by generating an AST which calls the `map/foreach` method on the same prefix object (lines XX-YY).

The next important step is the transformation of the passed closure AST into a local free method AST (lines XX-YY). For example, the passed closure “`x => x + 1`” would be transformed to “`def local1(x$1: Int): Int = x$1.+(1);`”. A closure like “`x => x + y`” where `y` is defined in the local scope would be transformed to something like “`def local1(x: Int): Int = x.+(TestMacroMapObject.this.y)`”. Also, a closure like “`println("hi"); x => x + 1`” would be transformed to “`println("hi"); def local1(x: Int): Int = x.+(1)`”

Subsection: Transformation Choice and Pattern

The transformation choice depends primarily upon the prefix’s type. Right now, there are three different transformation paths for different kinds of prefixes. We will check them in the following subsections but all of them are based on the same core transformation pattern. In pseudocode that structure could be described as:

```
reify val local_method = local_method_AST.splice // we saw it in the previous subsection
val inputCollection = prefix_AST.splice var outputCollectionBuf = output_collection_buffer_AST.splice
// will hold or generate the new collection while (prefix.hasNext()) val elem = prefix.getNext()
val result = local_method(elem) outputCollectionBuf.add(result) outputCollectionBuf.generate()
```

The reasons we need different implementations for different kinds of prefixes and target collections are: - Each kind has different API, e.g., different supported methods - Different implementation logic is needed for each kind in order to produce faster code, by exploiting each kinds’ specific characteristics since the same methods might take different time on different collections.

Since we are more interested in the sequences collections, as they are more common, trait `Seq` has two subtraits `LinearSeq`, and `IndexedSeq`. These do not add any new operations, but each offers different performance characteristics: A linear sequence has efficient head and tail operations, whereas an indexed sequence has efficient apply, length, and (if mutable) update operations.

Here, also, we see how important reifying and splicing operations are. The mentioned ASTs that we splice inside the `reify` above are constructed with one of these three ways depending on the occasion: - we get them directly from the FastTrack’s pattern matching - we get them by using the “`Symbol`”s and “`Type`”s APIs - we construct them manually.

Whatever AST is returned from the reify will eventually replace the macro application's AST in the first place.

Subsection: Handling `map`'s idiosyncrasies

In contrast to most of the other collections operations, a `map` operation can generate a collection that has the same type constructor as the prefix collection but possibly with a different element type, or even a entirely different collection type. As an example of the former case, if `f` is a function from `String` to `Int`, and `xs` is a `List[String]`, then `xs map f` gives a `List[Int]`. Likewise, if `ys` is an `Array[String]`, then `ys map f` gives a `Array[Int]`. As an example of the later case, `"Map("a" -> 1, "b" -> 2) map case (x, y) => y "` returns `List(1, 2)` of type `scala.collection.immutable.Iterable[Int]`, which is different from the prefix's `Map` type. The upstream declosurify project could handle partially only the first case. FT-DECLOSURIFY can handle both of the cases successfully due to the introduction of the implicit `CanBuildFrom` object, making it more general purpose.

Subsection: `Array`, `scala.collection.mutable.ArrayOps` and `scala.collection.mutable.IndexedSeq` Transformation

`scala.Array` and all collections that implement the `scala.collection.mutable.ArrayOps` and `scala.collection.mutable.IndexedSeq` traits, like `scala.collection.mutable.ArraySeq`, `scala.collection.mutable.ArrayBuffer`, share the same transformation, since all of them are indexed **mutable** sequences. The exact transformation is defined on lines XX-YY. As an example, any `macroMap` applications on `Arrays` like:

```
Array(1, 2, 3).macroMap(_ + 1)
```

will expand to:

```
def local1(x$1: Int): Int = x$1.+(1); val xs: Array[Int] = scala.this.Predef.intArrayOps(scala.Array.apply(
scala.this.Predef.wrapIntArray(Array[Int](2, 3))).repr(); var buf: Array[Int] = new Array[Int](xs.length());
var i: Int = 0; while(i.<(xs.length())) buf.update(i, local1(xs.apply(i))); // buf(i) = (local1(r(i))); i
= i.+(1) ; buf
```

Exactly the same transformation applies for all the other collections that fall into this category. The reasons we chose this transformation path for this category are: - only this category's collections accept a length in the constructor - the length method here takes constant time - the element selection through `xs.apply(i)` takes constant time - the update method is supported, since all collections of this category are mutable, and takes constant time

Also, all of this category's collections can only be mutated through the update method (the assignment operator) on a specific index, with one exception, the `scala.collection.mutable.ArrayBuffer`.

ArrayBuffer supports mutating methods for appending/removing elements which could affect the semantics of `macroMap` in case we mutate the underlying collection in the function we pass in `macroMap`. Our transformation seems to behave in the same way as the default `map` method. For example both “`buf map x => buf += x; x+1` ” and “`buf macroMap x => buf += x; x+1` ” return `ArrayBuffer(2, 3, 4)`. So, we assume the `macroMap` keeps the same semantics on the `ArrayBuffer` as the original `map` method.

In the next chapter will see how much faster this version is compared to the default `map` on `Arrays`.

Subsection: `scala.collection.LinearSeq` Transformation

This category includes `scala.collection.immutable.List`, `scala.collection.immutable.Stream`, `scala.collection.immutable.Queue`, `scala.collection.immutable.Stack`, `scala.collection.mutable.MutableList`, `scala.collection.mutable.MutableList`, `scala.collection.mutable.LinkedList`, `scala.collection.mutable.DoublyLinkedList`. All of these collections are linear. For example, in this category’s transformation we couldn’t use the `apply` (for indexing) or the `length` method because both of them take linear time, instead of constant. The transformation is defined on lines XX-YY. As an example, any `macroMap` applications on `scala.collection.immutable.List` like:

```
List(1, 2, 3).macroMap(_ + 1)
```

will be replaced from:

```
private def local1(x$1: Int): Int = x$1.+(1); private def builder1: scala.collection.mutable.Builder[Int,List[Int]] = val b: scala.collection.mutable.Builder[Int,List[Int]] = immutable.this.List.newBuilder(); b.sizeHint(immutable.this.List.newBuilder().sizeHint(1, 2, 3)); b ; val buf = builder1(); var these = immutable.this.List.newBuilder().newBuilder(); while(!these.isEmpty) buf.+=(local1(these.head)); these = these.tail ; buf.result
```

This category uses this transformation because: - `isEmpty` takes constant time - `head` takes constant time - `tail` takes constant time

In the next chapter will see how much faster this version is compared to the default `map` on `Lists`.

Subsection: `scala.collection.Traversable` Transformation

This category includes all collections that don’t fit in any of the previous categories. Roughly it includes all kinds of `Sets`, `Maps`, `Buffers` and `immutable.IndexedSeqs`. Obviously, since it includes such a broad range of collections we are allowed to use only methods that are supported from all collections, meaning we cannot use the `length`, `apply`, `head`, `tail` methods of the previous transformations. The transformation is defined on lines XX-YY. As an example, any `macroMap` applications on `scala.collection.immutable.Set` like:

`Set(1, 2, 3).macroMap(_ + 1)`

will expand to:

```
private def local1(x$1: Int): Int = x$1.+(1); private def builder1: scala.collection.mutable.Builder[Int, scala.collection.mutable.Set[Int]] = scala.collection.mutable.Builder[Int, scala.collection.mutable.Set[Int]](scala.collection.mutable.Set.empty[Int]);
= val b: scala.collection.mutable.Builder[Int, scala.collection.mutable.Set[Int]] = builder1; b.sizeHint(r); b ; val buf = builder1(); val it = r.toIterator; while(it.hasNext) buf.+=(local1(it.next));
; buf.result
```

Here, we use the `iterator` method to generate a prefix's iterator object. All collections have an `toIterator` method since all of them implement the `scala.collection.Iterable` trait. Each collection implements it using its `scala.collection.Traversable`'s `foreach` method, which is implemented according to each collection's special characteristics in order to be more efficient.

For the collections that implement the `immutable.IndexedSeq` trait, like `Vector`, we could have used a modified version of the mutable indexed sequences transformation where we would use a builder object to build the target collection, since we cannot mutate it in place. Interestingly, our experiments showed us that solution was slower than the current one.

Chapter 4

Safe Publication

Fixme
Fatal:
Replace
me

Chapter 5

Experimental Results

Fixme
Fatal:
Replace
me

Chapter 6

Related Work

Fixme
Fatal:
Replace
me

Chapter 7

Conclusions

Fixme
Fatal:
Replace
me

Acronyms and Abbreviations

Abbreviation	Full Name
--------------	-----------

Appendix A

Escape Analysis Code

Fixme
Fatal:
Add me