

#### NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

#### SCHOOL OF SCIENCES

## DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS

#### POSTGRADUATE STUDIES PROGRAM

#### **MASTER THESIS**

## Faster Scala Collections with Compile-Time Reflection

**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

#### ΔΙΠΛΩΜΑΤΙΚΉ ΕΡΓΑΣΙΑ

## Γρήγορες Δομές Δεδομένων στη Γλώσσα Scala με Στατική Ανάκλαση

Γεώργιος Κόλλιας

Επιβλέπων: Γιάννης Σμαραγδάκης, Αναπληρωτής Καθηγητής ΕΚΠΑ

**AOHNA** 

**MAIOΣ 2013** 

#### **MASTER THESIS**

### Faster Scala Collections with Compile-Time Reflection

## **Georgios Kollias**

RN: M1049

#### **SUPERVISOR:**

Yannis Smaragdakis, Associate Professor NKUA

#### THESIS COMMITTEE:

Yannis Smaragdakis, Associate Professor NKUA Panos Rondogiannis, Associate Professor NKUA

#### ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

### Γρήγορες Δομές Δεδομένων στη Γλώσσα Scala με Στατική Ανάκλαση

Γεώργιος Κόλλιας ΑΜ: Μ1049

#### ΕΠΙΒΛΕΠΩΝ:

Γιάννης Σμαραγδάκης, Αναπληρωτής Καθηγητής ΕΚΠΑ

#### ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:

**Γιάννης Σμαραγδάκης**, Αναπληρωτής Καθηγητής ΕΚΠΑ **Παναγιώτης Ροντογιάννης**, Αναπληρωτής Καθηγητής ΕΚΠΑ

Περίληψη

Περιγράφουμε την υλοποίηση συγκεκριμένων μεθόδων (map και foreach) των δομών δε-

δομένων στη βιβλιοθήκη της γλώσσας Scala, με χρήση λειτουργιών στατικής ανάκλασης. Ο

σκοπός είναι η δημιουργία γρηγορότερων δομών μέσω ενσωμάτωσης και βελτιστοποίησης της

λειτουργίας στο σημείο κλήσης. Η λειτουργικότητα αυτή περιλαμβάνεται στη βιβλιοθήκη της

γλώσσας, έτσι ώστε ταχύτερες λειτουργίες μπορούν να χρησιμοποιηθούν σε όλους τους τύπους

δομών της γλώσσας (π.χ. λίστες, πίνακες, κτλ.) χωρίς ανάγκη ορισμού νέων τύπων. Ο

μηχανισμός μας έχει υλοποιηθεί κατέυθείαν στη βασική βιβλιοθήκη και το μεταγλωττιστή της

γλώσσας. Τα αποτελέσματα είναι ενθαρρυντικά, με πειράματα να εμφανίζουν ταχύτητα βελτιωμένη

κατά 40%.

**ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ:** Στατικός Μεταπρογραμματισμός

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ:** Scala, μεταπρογραμματισμός, δομές δεδομένων, μεταγλωττιστές, μακροεν-

τολές

**Abstract** 

We describe the implementation of specific Scala collections operations (currently the map

and foreach methods) using the Scala 2.10 compile-time reflection facilities. The primary

motivation for this work is to create faster collections by inlining operations at the call site.

The functionality is available at the standard Scala library level, so that our optimized oper-

ations can be used on all plain Scala collection types (e.g., List, Array, etc.) without the

need of creating new specialized types. Our mechanism is implemented directly inside the

Scala standard library and by modifying the default compiler. The results are encouraging

since benchmarks show a 40% speedup.

**SUBJECT AREA:** Compile-Time Metaprogramming

**KEYWORDS:** Scala, metaprogramming, collections, compilers, macros

# Acknowledgements

FiXme
Fatal:
Add me

### **Contents**

1	Intr	oducti	on	13	
2	Bac	kgroun	ıd	16	
	2.1	Scala	Collections Overview	16	
	2.2	Scala	Compile-Time Reflection Overview	17	
3	Our	Appro	ach: ft-declosurify	21	
	3.1	FT-DEC	CLOSURIFY Overview	21	
	3.2	FT-DEC	CLOSURIFY Implementation Specifics	24	
		3.2.1	macroMap/macroForeach definitions	24	
		3.2.2	FT-DECLOSURIFY in the Scala standard library	25	
		3.2.3	Transformation Method Interface	27	
		3.2.4	Transformation Requirements	28	
		3.2.5	Transformation Choice and Idiosyncrasies	29	
		3.2.6	Array, scala.collection.mutable.ArraysOps and scala.collection	on.	
			mutable.IndexedSeq Transformation	30	
		3.2.7	scala.collection.LinearSeq Transformation	32	
		3.2.8	scala.collection.Traversable Transformation	34	
4	Experimental Results				
	4.1	Setup		36	
	4.2	Evalua	ation	36	
5	Rela	ated W	ork	40	
6	Con	clusio	ns	41	
Ac	rony	ms and	d Abbreviations	42	
Αŗ	pend	lices		43	
A	map	Infix	Code???	43	

# List of Figures

1.1	Example of naive map in Java	14
1.2	Example of generic map in Java	14
2.1	All scala.collection collections	17
3.1	<pre>scalac's internal representation of List(1, 2, 3).map(_ + 1)</pre>	21
3.2	Scala's default map implementation	22
3.3	List(1, 2, 3).macroMap(_ + 1) example transformation	23
4.1	Benchmarks of mutable indexed sequences representatives	37
4.2	List benchmark	38
4.3	Benchmarks of Traversable representatives	39

## List of Tables

# **List of Corrections**

Fatal: Add me	7
Fatal: Replace me	12
Fatal: add abbr	13
Fatal: cite	13
Fatal: cite http://www.infoq.com/news/2011/11/yammer-scala	15
Fatal: cite	16
Fatal: cite	16
Fatal: cite	16
Fatal: date	17
Fatal: cite Template Haskell	18
Fatal: cite Nemerle	18
Fatal: abbr	19
Fatal: cite 19, 8	19
Fatal: cite implicit conversions	25
Fatal: abbr	26
Fatal: cite	26
Fatal: cite adrian moors phd	27
Fatal: defined on lines XX-YY	30
Fatal: defined on lines XX-YY	32
Fatal: defined on lines XX-YY	34
Fatal: cite	36
Fatal: appendix	36
Fatal: cite	38
Fatal: cite miniboxing and I.Dragos' PhD	38
Fatal: Replace me	40
Fatal: Replace me	41
Fatal: Add me	43

# **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 FiXme 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.

Fatal: add abbr

Section: The Problem and Our Approach

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

Fatal: cite

"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. Figure 1.1 shows 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.

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

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

Figure 1.1: Example of naive map in Java

```
// A sample iterator function
void map( CallableOneArg fcn1arg, long[] dst, long[] src) {
   for(int i=0; i < dst.len; i++)
      dst[i] = fcn1arg.call(src[i]);
}</pre>
```

Figure 1.2: Example of generic map in Java

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. Figure 1.2 uses Callable.

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 fcnlarg.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

fxfatalcite github url.

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

ft-declosurify defines two methods, macroMap and macroForeach, that can be used from all the Scala standard libary 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.

Fatal: cite http://ww scala

FiXme

## Chapter 2

# **Background**

Scala is a relatively new statically typed programming language that tries to unify the objectoriented 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

Fatal: cite **FiXme** 

#### **Scala Collections Overview** 2.1

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 subpackages 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

16 Georgios Kollias

**FiXme** 

Fatal: cite

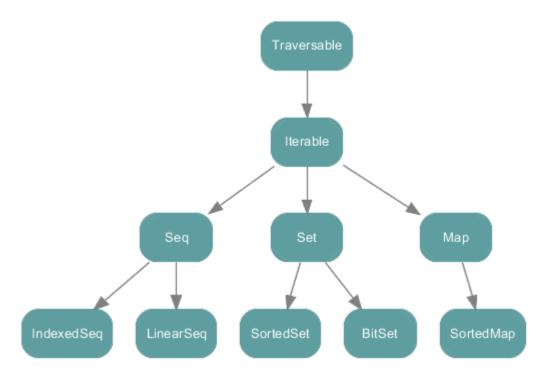


Figure 2.1: Basic collections in scala.collection

the mutable collections in package scala.collection.mutable typically add some sideeffecting 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 FiXme

Fatal: date

<sup>&</sup>lt;sup>1</sup>Figure 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.

FiXme
Fatal:
cite
Template
Haskell
FiXme
Fatal:
cite
Nemerle

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.

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 Fixme 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:

```
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 scalar 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 scalar to: (- + 1)

We can see that scalar 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-

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

<sup>&</sup>lt;sup>1</sup>The code listing below as well as most of the following listings are slightly abbreviated for readability reasons

```
def map[B, That](f: A => B)(implicit bf: CanBuildFrom[Repr, B, That]): That =

def builder = {
    val b = bf(repr)
    b.sizeHint(this)
    b
    val b = builder
    for (x <- this) b += f(x)
    b.result
}</pre>
```

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 al = new \$anonfun(); al(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  $new \quad sanonfun()$  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 1.2. 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 exlained 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 repsective 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

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

Initally, it seems that we fall back to The Problem without having any advantage, since <code>myf.apply</code> seems to be a megamorphic virtual call again. But, in a program, all <code>macroMap</code> 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 <code>myf.apply</code>, 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, making it inlining friendly.
- Having a fixed built-in implementation, applying eager compile-time inlining.

## 3.2 FT-DECLOSURIFY Implementation Specifics

In this section, we will see the textscft-declosurify implementation in more detail.

## 3.2.1 macroMap/macroForeach definitions

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:

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 depeding on their type parameters. Generally, builders makes the creation of a new collection out of existing ones easier and more maintable. 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.

### 3.2.2 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.TraversableOnce trait and Arrays, through Scala's 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

Fixme
Fatal:
cite
implicit
conversions

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 appication's call-site context and all of its arguments' Abstract Syntax Trees (ASTs), through pattern FiXme 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)+ 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.

Fatal:

abbr

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) Type Tag, that store the type parameters' full types on each call and FiXme 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 us 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.

Fatal: cite

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.

#### 3.2.3 **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) Type Tag'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
- bfTree: the AST of the implicit scala.collection.generic.CanBuildFrom object FiXme 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.

Fatal: cite adrian moors phd

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-depedent 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.Types that follow.

### 3.2.4 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. 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:

- 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:

```
def mkFallbackImpl: c.Tree = {
   val name: TermName = if (isForeach) "foreach" else "map"
   val pre = c.prefix.tree
   val fallbacktree = Apply(Select(pre, name), f0.tree :: Nil)
   fallbacktree
}
```

The next important step is the transformation of the passed closure AST into a local free method AST:

```
def functionToLocalMethod(fnTree: Function): DefDef = {
   val Function(fparams, fbody) = fnTree
   val frestpe = fbody.tpe
   val fsyms = fparams map (_.symbol)
```

For example, the passed closure  $x \Rightarrow x + 1$  would be transformed to

```
def local1(x$1: Int): Int = x$1.+(1);
```

. A closure like  $x \Rightarrow x + y$  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");  $x \Rightarrow x + 1$  would be transformed to

```
println("hi");
def local1(x: Int): Int = x.+(1);
```

### 3.2.5 Transformation Choice and Idiosyncrasies

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 generate similar code with the original Scala map implementation, although they use more low-level constructs.

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.

For example, 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 and Type APIs.
- we construct them manually.

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

Finally, we should also mention that 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 a more general purpose transformation tool.

# 3.2.6 Array, scala.collection.mutable.ArraysOps 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.StringBuilder, scala.collection.mutable.ArrayBuffer, share the same transformation, since all of them are indexed *mutable* sequences. The exact transformation is:

FiXme Fatal: defined

30 on lines

```
def mkMutIndexed[Prefix] (prefixTree: Tree): c.Tree = {
                  = c.Expr[Prefix] (prefixTree)
    val prefix
                = c.Expr[Int] ('xs dot 'length) // might be array or indexedseq
    val len
                  = c.Expr[Unit](('buf dot 'update)('i, closure('xs('i))))
    val call
    def mkResult = c.Expr[Nothing](if (isForeach) mkUnit else 'buf)
    val arrCons = Apply(Select(New(TypeTree(outCollTpe)), nme.CONSTRUCTOR), List
        (('xs dot 'length).lhs))
    val builderVal = c.Expr[Prefix] (arrCons)
    reify {
10
      closureDef.splice
11
      val xs = prefix.splice
      var buf = builderVal.splice
13
      var i = 0
14
      while (i < len.splice) {</pre>
        call.splice
16
        i += 1
17
18
      mkResult.splice
19
    }.tree
21 }
```

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(1,
  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
  buf</pre>
```

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.

### 3.2.7 scala.collection.LinearSeq Transformation

This category includes scala.collection.{immutable.List, immutable.Stream, immutable.Queue, immutable.Stack, mutable.MutableList, mutable.MutableList, mutable.LinkedList, mutable.LinkedList, mutable.DoubleLinkedList}. 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:

FiXme
Fatal:
defined
on lines
XX-YY

```
def mkLinear(prefixTree: Tree): c.Tree = {
    val prefix = c.Expr[Lin[A]] (prefixTree)
    val call
              = mkCall('these dot 'head)
    reify {
      closureDef.splice
      builderDef.splice
      builderVal.splice
      var these = prefix.splice
      while (!these.isEmpty) {
10
        call.splice
        these = these.tail
      mkResult.splice
    }.tree
15
16 }
```

As an example, any macroMap applications on scala.collection.immutable.List like:

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

#### will be replaced from:

```
{
1
    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.canBuildFrom[Int].apply();
      b.sizeHint(immutable.this.List.apply[Int](1, 2, 3));
      b
    };
    val buf = builder1();
    var these = immutable.this.List.apply[Int](1, 2, 3);
10
    while(!these.isEmpty) {
      buf.+=(local1(these.head));
12
      these = these.tail
13
    };
14
    buf.result
15
```

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.

#### 3.2.8 scala.collection.Traversable Transformation

This category includes all colletions 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:

FiXme
Fatal:
defined
on lines
XX-YY

```
def mkTraversable(prefixTree: Tree): c.Tree = {
   val prefix = c.Expr[Traversable[A]](prefixTree)

  val call = mkCall('it dot 'next)

  reify {
    closureDef.splice
    builderDef.splice
    builderVal.splice
   val it = prefix.splice.toIterator
   while (it.hasNext)
        call.splice

mkResult.splice

// MkResult.splice
// Traversable[A]](prefixTree)
// Next)
// Colored in the prefix in the pre
```

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:
```

```
4 scala.collection.mutable.Builder[Int,scala.collection.immutable.Set[Int]] = {
      val b:
  scala.collection.mutable.Builder[Int,scala.collection.immutable.Set[Int]] =
  immutable.this.Set.canBuildFrom[Int].apply();
      b.sizeHint(r);
      b
9
    };
11
    val buf = builder1();
    val it = r.toIterator;
12
    while(it.hasNext) {
13
      buf.+=(local1(it.next));
14
    } ;
15
    buf.result
16
17
  }
```

Here, we use the iterator method to generate a prefix's iterator object. All collections have an tolterator 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

# **Experimental Results**

In this chapter we will benchmark all the categories we explored in the previous chapter against their default Scala counterparts. We choose a few representative collections for each category in order to see how the different transformation strategies and the different collections characteristics affect the speedups.

#### 4.1 Setup

We used the ScalaMeter microbenchmarking and performance regression testing framework Fixme for the JVM platform. We use the same benchmark template for all the different categories. The code for the Array benchmarking is in Appendix X. In short, each benchmark compares the macroMap's speed against map's speed on different sizes of the same collection applying the same closure, the successor function.

Fatal: cite **FiXme** Fatal: appendix

The benchmarks run on a Linux 64-bit machine with an Intel Core i7-2720QM 8-core CPU and 8GB RAM.

In order to make the benchmarking process more stable, reproducible and reliable we have tweaked the benchmark template with the following parameters:

- each measurement of map/macroMap on a specific collection size is run 25 times successively
- the 25 measurents are divided on 5 separate JVM instances
- every 2 measurements the current collection is reinstantiated
- every 2 measurements a full GC cycle is forced

#### **Evaluation** 4.2

In Figure 4.1 we see the benchmark results for scala.collection.mutable.ArraySeq and Array respectively, as representatives of the mutable indexed sequences category, i.e., subtypes of scala.collection.mutable.IndexedSeq and Array.

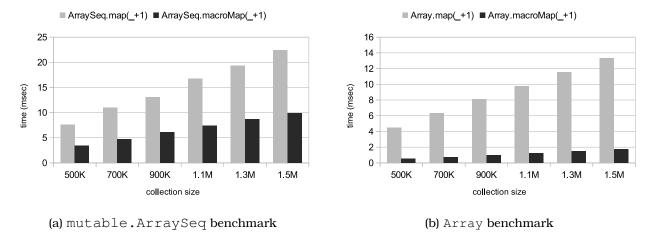


Figure 4.1: Benchmarks of mutable indexed sequences representatives

The speedup on Array is impressive and the highest one among all the other benchmarked collections. In this series of benchmarks, all collections hold integers and the supplied closure returns an integer too (the successor integer). Scala has only one integer type, scala.Int, which is similar to Java's boxed java.lang.Integer. scalac tries to use Java's respective primitive types under the hood, whenever possible, for optimization reasons. More specifically, after scalac's erasure phase, a scala.Int will become either a int or a java.lang.Integer. For example, in our benchmarks, the closure function will become a local method like:

```
def local1\$1(x\$1: int): int = x\$1.+(1);
```

This method deals solely with Java's primitive integers under the hood. On the other hand operations on ArraySeq, List and most of the collections except Array, deal with java.lang.Integer. So, for example, in an ArraySeq transformation the target collection is constructed like this:

```
buf.update(i, scala.Int.box(local1\$1(scala.Int.unbox(xs.apply(i))
```

#### while on Arrays we have:

```
buf.update(i, local1\$1(xs.apply(i)))
```

This is due the way Arrays are constructed. In particular, when we ask for an Array[scala.Int] in Scala, the compiler inspects it and creates an Array[int] automatically and, as a consequence, all of its methods operate upon primitive integers. We can

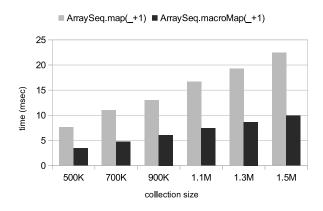


Figure 4.2: List benchmark

see that avoiding the boxing/unboxing operations can give us a huge performance boost. For that reason, Scala already provides the @specialized annotation for creating specialized classes/collections and, more generally, specialization is a hot research topic in the Scala ecosystem .

In Figure 4.2 we see the benchmark results for scala.collection.immutable.List, representative of the linear sequences category, i.e., subtypes of scala.collection.LinearSeq.

In Figure 4.3, we see the benchmark results for scala.collection.immutable.Set and scala.collection.immutable.Vector, representatives of the general category of traversables, i.e., subtypes of Traversable, which includes all collections.

FiXme
Fatal:
cite
FiXme
Fatal:
cite miniboxing
and
I.Dragos'
PhD

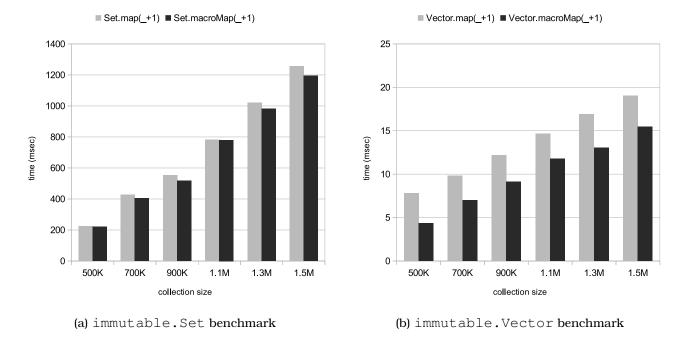


Figure 4.3: Benchmarks of Traversable representatives

# Chapter 5

# **Related Work**

FiXme
Fatal:
Replace
me

# Chapter 6

# **Conclusions**

FiXme
Fatal:
Replace
me

# **Acronyms and Abbreviations**

Abbreviation Full Name

# Appendix A

# mapInfix Code???

FiXme Fatal: Add me