The Plaid Language: Typed Core Specification

version 0.4.0

Jonathan Aldrich **Karl Naden** **Sven Stork**
Joshua Sunshine **Robert Bocchino**

December 2011
CMU-CS-11-XXX

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

This document defines the core of the Plaid language, including its initial type system

This research was supported by DARPA grant HR00110710019, CMU|Portugal Aeminium grant CMU-PT/SE/0038/2008, NSF grants CCF-0811592 and CCF-1116907, and grant #1019343 to the Computing Research Association for the CIFellows Project.

Keywords: programming language, typestate, Plaid, gradual typing, permissions

1 Conventions

This document uses the same grammar definition conventions as the Java Language Specification, Third Edition (JLS) [Gosling et al., 2005]. Those conventions are described in chapter 2 of the JLS and are not repeated here.

2 Lexical Structure

The lexical structure of Plaid matches that of Java, as defined in chapter 3 of the JLS. Specifically:

- Plaid uses the same definitions of line terminators as Java (JLS section 3.4), the same input elements and tokens (JLS section 3.5) except for a different keyword list, and the same definition of whitespace (JLS section 3.6).
- Plaid uses the same definition of comments as Java (JLS section 3.7).
- Plaid uses the same definition of identifiers as Java (JLS section 3.8).
- Plaid literals are based on Java literals (JLS section 3.10), but there are several substantial differences. Plaid string literals are the same as Java string literals. Plaid integer literals are of arbitrary length. Plaid rational literals are like Java double literals but are of arbitrary length. Furthermore, Boolean objects named `true` and `false` exist in the standard library, but unlike in Java these are not keywords in Plaid. No other Java literals are currently supported, but future versions of Plaid will support all Java literals except the null literal.
- Plaid uses the same definition of Separators as Java (JLS section 3.11).

2.1 Keywords

The following character sequences, formed from ASCII letters, are reserved for use as *keywords* and cannot be used as identifiers:

Keyword: one of

atomic	**callonce**	**case**	**default**	**dyn**
dynamic	**exclusive**	**fn**	**freeze**	**full**
group	**immutable**	**import**	**local**	**match**
method	**mutable**	**new**	**none**	**of**
override	**package**	**pure**	**readonly**	**remove**
rename	**requires**	**shared**	**split**	**state**
stateval	**this**	**top**	**type**	**unique**
unpack	**val**	**var**	**void**	**with**

2.2 Operators

We first define operator characters as follows:

OperatorChar: one of

`= < > ! ~ ? : & | + - * / ^ %`

Now an operator is a sequence of operator characters:

Operator:

OperatorChar

OperatorChar Operator

The exception to the grammar above is that the character sequences =, =>, <<-, >> and <- have other meanings in the language and may not be used as operators. Furthermore, operators containing the comment sequences `/*` or `//` may not be used as operators.

3 Statements and Expressions

3.1 Exceptions

Several locations in this document refer to an exception being thrown. The semantics of an exception being thrown is that the application halts with a run-time error. Future versions of this document will define facilities for propagating and catching exceptions.

3.2 Statements

Stmt:

Expr

VarDecl

StateValDecl

VarDecl:

Specifier [Type] Identifier = Expr

StateValDecl:

`stateval` *Identifier StateBinding*

Specifier:

`val`

`var`

Statements are either expressions, or variable declarations. A variable declaration must include an initial value. Object variables are declared with the **`val`** or **`var`** keyword; the former indicates a final let binding, whereas the latter indicates a assignable variable that can be updated. State variables are declared with the **`stateval`** keyword.

An optional type may be given for variable declarations. If the type is omitted for a **`val`** declaration, then it has the structure of the initializing expression and the default permission for that structure. If no type is given for a **`var`** declaration, the variable is has type **`dynamic`**.

Statements evaluate to values, based on the expression in the statement or the value of the initializer for the variable. The last statement in a sequence is used for the return value of a method or the result of a block.

3.3 Expressions

Expr:

`fn` *[MetaArgsSpec]([Args]) [[Args]] => Expr*

Expr1

A first-class function includes standard polymorphic and parameter argument declarations. The optional arguments surrounded by [] specify environment variables that are captured by the lambda at its declaration. The syntax of polymorphic parameters are described on Section 4. Polymorphic parameters specify which data group permission a functions requires and allow the function to be generic on which data groups it operates (similar to generic methods on Java).

Expr1:

[SimpleExpr .] Identifier = Expr

SimpleExpr <− State

SimpleExpr <<− State

`atomic` *MetaArgs* BlockExpr

`split` *MetaArgs* BlockExpr

`unpack` BlockExpr

`match` (*InfixExpr*) { *{*CaseClause*}* }

InfixExpr

The assignment form is for fields or for already-declared local variables, which must have been declared using **`var`**.

The state change operator <− modifies the object to the left of the arrow as follows:

- All tags on the right are added to the object. Old tags are kept unless they are inconsistent with the new tags, i.e. the old tag and new tag are (transitively) different cases of the same state.

- All members that were declared in tags being removed, are removed from the object.
- All members on the right are added to the object. All old members on the left that are not explicitly removed according to the bullet above are retained.
- Futher details on the semantics of state change can be found in Sunshine et al. [2011].

The replacement operator <<− removes all tags and members from the object on the left and adds all tags and members of the state on the right.

The type of either state change operations is **`void`**.

This paragraph provides a short overview of the Æminium-specific expressions. For a full definition of the semantics refer to [Stork et al., 2009, 2010]. The **`atomic`** expression provides a safe access environment to all shared objects which belong to the data groups mentioned by the **`atomic`** block. The *MetaArgs* describes which data groups the atomic block guarantees mutual exclusive access. Each *MetaArg* must refer to a valid data group (i.e., a concrete declared data group or a group parameter). The **`split`** executes all statements of its body concurrently. To allow parallel access to shared data the **`split`** block will split the declared data group permissions into shared permissions, one for each statement. **`unpack`** is used to trade the group/access permission to the specified object to gain access to the inner/nested groups declared inside the object.

The **`match`** expression matches an input expression to one of several cases using the *CaseClause* construct defined below. The overall match expression evaluates to whatever value the chosen case body evaluates to.

CaseClause:

`case` *Pattern BlockExpr*

`default` *BlockExpr*

Pattern:

QualifiedIdentifier

QualifiedIdentifier:

Identifier { . Identifier}

The value is matched against each of the cases in order. For the first case that matches, the corresponding expression list is evaluated. If no pattern matches, an exception is thrown.

A **`case`** tests the value's tags against the *QualifiedIdentifier* given by the specified pattern. The match succeeds if one of the tags of the matched value is equal to the tag *QualifiedIdentifier*, or if one of the tags of the matched value was declared in a state that is a transitive **`case of`** the *QualifiedIdentifier* specified.

The *QualifiedIdentifier* must resolve to a state declared with the **`state`** keyword; otherwise, an exception is thrown.

For the **`default`** case, the match always succeeds. If there is a **`default`** case, it must be the last one in the match expression.

InfixExpr:

SimpleExpr

CastExpr

InfixExpr IdentifierOrOperator InfixExpr

IdentifierOrOperator:

Identifier | *Operator*

CastExpr:

*SimpleExpr [***as** *Type]*

The operators defined in Java have the same precedence in Plaid as they do in Java, except the ternary operator and right shift operators, which are unsupported. Identifiers as well as symbolic operators can be used as infix operators; both are treated as method calls on the object on the left of the operator. Non-Java operators and identifiers used as infix operators have a precedence above assignment and state change, and below all other operators.

Cast expressions assert that a variable has a given type, and also assert the relevant permission for that variable. These casts are trusted by the typechecker, but unchecked. A program that executes an invalid cast may fail at any point later in the program's execution.

SimpleExpr:

BlockExpr

new *State*

SimpleExpr2

The **new** statement creates an object initialized according to the *State* specification given (defined below).

BlockExpr:

{ *[StmtListSemi]* }

StmtListSemi:

Stmt { ; *Stmt} [*;*]*

Block expressions have a semicolon-separated list of statements, with an optional semicolon at the end. The statement list evaluates to the value given by the last statement in the list.

SimpleExpr2:

SimpleExpr1

SimpleExpr2 BlockExpr

To enable control structures with a natural, Java-like syntax, we allow a function to be invoked passing a block expression as an argument. The block expression in this case is a zero-argument lambda.

SimpleExpr1:

Literal

Identifier

this

(*Expr*)

SimpleExpr1 . *Identifier*

SimpleExpr1 . **new**

SimpleExpr1 [MetaArgs] ArgumentExpr

ExprList:

Expr { , Expr}

ArgumentExpr:

(*[ExprList]*)

this represents the receiver of a method call as in Java and is bound in method bodies declared as members of states. Unlike Java, **this** is not bound in field initializers.

Expressions can appear within parentheses as a comma separated list representing a tuple.

Java constructors can be invoked by calling **new** on the Java class name.

Function and method invocation are handled uniformly by supplying the arguments as a tuple. Applications can be chained, supporting currying. Polymorphic arguments are specified at each call site as well.

4 Polymorphism

MetaArgsSpec:

< *MetaArgSpec [, MetaArgSpec]* >

MetaArgSpec:

group *[GroupPermission] Identifier*

GroupPermission:

exclusive

shared

protected

MetaArgs:

< *SimpleExpr1 [, SimpleExpr1]* >

Plaid supports polymorphism for data groups.[1] Plaid uses angle bracket to enclose polymorphic parameters and arguments (similar to Java's generics). A *MetaArgSpec* describes a single polymorphic formal parameter. At the moment Paid only supports only group parameters. A group parameter consists of the **`group`** keyword to identify this parameter as group parameter, and optional *GroupPermission* (only optional for state declarations) and the name of the parameter. For more information about data groups and group parameters refer to [Stork et al., 2009, 2010].

5 Declarations

DeclOrStateOp:

Decl

StateOp

Decl:

{ModifierOrDefaultPermission} **`state`** *Identifier [MetaArgs]*
*[***`case of`** *QualifiedIdentifier [MetaArgs]] [StateBinding] [;]*

{ModifierOrDefaultPermission} **`stateval`** *Identifier [MetaArgs]*
[StateBinding] [;]

{Modifier} MSpec ;

{Modifier} MSpec BlockExpr

{Modifier} FieldDecl ;

{Modifier} GroupDecl ;

`state` and **`stateval`** declarations specify the implementation of a state, as specified in the state definition. The **`state`** keyword means that this state is given its own *tag* that can be used to test whether objects are in that state. Only states declared with **`state`** can be given in a pattern for a case in a **`match`** statement.

The **`case of`** keyword assigns a superstate. States have all of the members of a superstate. Different cases of the same superstate are orthogonal; no object may ever be tagged with two cases of the same superstate.

The final two declarations are for method and field declarations. The method declaration has a method header and an optional method body. If the body is missing then the method is abstract and must be filled in by sub-states or when the state is instantiated.

Fields and state operators are discussed in more detail below.

[1] Extending polymorphism to types should be straight forward.

StateBinding:

= *State*

{ *{Decl}* }

State:

*StatePrim {***with** *StatePrim}*

StatePrim:

SimpleExpr1 [{ {DecOrStateOp} }]

{ *{Decl}* }

freeze *SimpleExpr1*

StateOp:

remove *Identifier* ;

rename *Identifier* **as** *Identifier* ;

A state is a composition of primitive states separated by the **with** keyword. Primitive states include literal blocks with a series of declarations and references to some previous object or state definition. In addition, objects can be transformed into primitive states with the **freeze** keyword.

Composition is in general symmetric, as in traits. It is an error if two states are composed with a member in common. The conflict can be resolved manually using state operators **remove** and **rename**, which respectively discard or change the name of a member in a state.

GroupDecl:

group *Identifier* = **new group** ;

FieldDecl:

ConcreteFieldDecl

AbstractFieldDecl

ConcreteFieldDecl:

*[Specifier] [*Type*] Identifier = Expr*

AbstractFieldDecl:

Specifier Identifier

*[*Specifier*] Type Identifier*

MSpec:

method *[Type] IdentifierOrOperator [MetaArgsSpec]* (*[Args]*) *[*[*Args*]*]*

Args:

[ArgSpec] Identifier { , [ArgSpec] Identifier}

ArgSpec:

Type [>> Type]

The *FieldDecl* form should be familiar from Java-like languages. If no expression is given then the field is abstract. When fields are first defined a specifier (**var** or **val**) must be given; later, when the field is overridden and given a concrete value, the specifier may be omitted. **var** fields are assignable, **val** fields are not.

If a type is missing and an expression is given for a **val** field, then the type of the field is inferred from the expression as in variable declaration statements. If the type is missing and either no expression is given or it is a **var** field, then the type is **dynamic**.

The method header *MSpec* also has a standard format (similar to functions). As in function declarations, programmers may optionally include types for any captured environment variables within []. For methods declared within states, the distinguished variable **this**, representing the receiver of the method, may appear in this list. If it does not, then the type of the receiver defaults to the structure representing the state the method is defined in with the default permission for that state. The receiver ends the method with the same type.

Each argument specification includes the required type at the time of the method call or function application. If the parameter ends the call with a different type this is indicated with a $>>$ and the resulting type. If no resulting type is specified then it defaults to the required type. If no argument specification is given for a variable, then its starting and ending type defaults to **dynamic**.

Modifier:

requires

override

DefaultPermission:

immutable

ModifierOrDefaultPermission:

Modifier

DefaultPermission

override indicates that a method overrides a function of the same name during composition.

requires is similar to **abstract** in Java. However, things are more interesting in Plaid, because one can pass around an object that has abstract/required members. It is not necessary to use the **requires** modifier in state definitions; one can simply leave off the definition of a function. **requires** is necessary in types, however, to distinguish the presence vs. absence of a member in that type. Unlike in Java, methods may be called on an object that has a required member, but only if the type given to the method's receiver does not expect that member to be present.

immutable means a state is immutable and the permission of any fields, local variables, or parameters declared to have the structure represented by the state defaults to **immutable** when a permission is not specified. If the state is not immutable then the default permission is **unique**.

6 Types

Type:

void

*[***callonce***] LambdaStructure*

[Permission] NominalStructure

(*Type*)

Permission:

unique

SymmetricPermission

LocalPermission

SymmetricPermission:

shared <*SimpleExpr1*>

immutable

LocalPermission:

local *SymmetricPermission*

LambdaStructure:

[MetaParams] (*ArgSpecs*) -> *Type*

NominalStructure:

top

*QualifiedIdentifier [*MetaArgs*]*

ArgSpecs:

*ArgSpec { * ArgSpec }*

All types in Plaid include a permission and a structure. The most general type is **void** which represents the weakest permission, **none** (not otherwise expressible in the source), with the most general structure, **top**. References may be inferred to have the type **dynamic**. Uses of values with type **dynamic** are not statically guaranteed to be type-safe. An unsafe cast must appear in the source for a **dynamic** value to be used in statically typed code. Lambda types consist of a lambda structure with an optional lambda permission **callonce**. Normal functions can be called any number of times, while a **callonce** function can only be called once. All other types are written as an optional *Permission* and a structure.

A *LambdaStructure* represents a function structure. The environment included in the declaration of a lambda is not included in its type. However, if a function is specified to consume part of the permission to any of its environment variables, then it can only be typed at a **callonce** lambda type since the necessary environment permission will no longer be available after a call.

Consequently, once a **callonce** function has been called, its type becomes **void**. Formally, a function that accepts multiple arguments actually accepts an argument tuple, which is written with a $*$-separated list.

A *NominalStructure* represents the structure given by a declared state or the distinguished **top** structure, which is a superstructure of all structures. If the state represented by the *NominalStructure* is polymorphic, then *MetaArgs* must be provided.

If the permission for a nominal structure is not given, then a default is inferred. The **top** structure defaults to the **none** permission and a *NominalStructure* defaults to the **immutable** permission if it represents an **immutable** state and to **unique** otherwise.

The **unique** permission indicates that there are no usable aliases to the same object. There may be other references to the object with the **none** permission which does not allow the object to be used in any way.

A *SymmetricPermission* allow new aliases to be created with the same permission. **immutable** references cannot be used to update the object but can assume that no other references make changes. **shared** references can make changes, but must assume that other references may have changed the object.

A **local** permission gives the same abilities and guarantees as its underlying *SymmetricPermission*, but is restricted to local variables and parameters. A **local** reference cannot be assigned into a field. This restriction allows **local** permissions to be returned to their original location when their reference goes out of scope. This may allow the original reference to regain a stronger permission. For example, a **unique** reference used as a function parameter that requires and results in a **local** permission will still be **unique** after the function call. Refer to [Naden et al., 2012] for more information on **local** permissions.

7 Compilation Units

CompilationUnit:

package *QualifiedIdentifier* **;** *{Import} {Decl}*

A compilation unit is made up of a (required) package clause followed by a sequence of declarations.

7.1 Imports

Import:

import *QualifiedIdentifier [DotStar]* **;**

DotStar:

. *

An import statement imports a qualified name into the current scope so it can be referred to by the last identifier in the qualified name. If the import ends in .*, then all the members of the given *Name* are imported into the current scope.

As in Java, importing the same simple name twice is an error unless the fully qualified name is the same. Importing a specific simple name always overrides importing all elements of a package where that name is defined, regardless of which definition goes first. In general, Plaid follows the Java Language Specification section 7.5.

7.2 Java Interoperability

Accessing Java from Plaid. Any java package, class, or class member can be referred to via a qualified name. Imported name(s) can include a package, class, or class member from Java. Instances of a Java class C may be created by invoking C.new(...) and passing appropriate arguments for one of the constructors of class C. A static method m of C may be invoked with the syntax C.m(...). An instance methods of a Java object o may be invoked with the syntax o.m(...). Arguments passed to calls of Java constructors and methods may be Java objects. Plaid integers, strings, and booleans are converted to appropriate Java primitive, String, and numeric object types (e.g. java.lang.Integer) depending on the declared type of the method's formal parameters. If a Java method takes an Object or plaid.runtime.PlaidObject as an argument, then a Plaid object can be passed to it, allowing Java code to access Plaid objects.

Implementing Java Interfaces. A Plaid state can be declared to be a case of a Java interface. In that case, any **new** expression that creates an object with that state will generate a Plaid object that extends the appropriate Java interface. The Plaid object may then be passed to a Java method that takes the interface type as an argument. Methods of the interface that are invoked by Java are converted into calls to Plaid methods of the same name and arguments, as described immediately below.

Accessing Plaid from Java. Java code may invoke methods of Plaid objects when those objects implement Java interfaces, as decribed above, or reflectively through the plaid.runtime.PlaidObject interface. When calling a Plaid method through this interface, Java objects of type Integer, String, Booleans, and other numeric objects are converted into the corresponding Plaid types. PlaidObjects and Java objects are passed through unchanged, and their methods may be invoked from Plaid in the usual way described above. The detailed interface of plaid.runtime.PlaidObject is specified in the javadoc for that interface.

7.3 File System Conventions

A compilation unit is stored in a file with extension .plaid. For each top-level declaration in the file, a Java class in the package declared is created with the name of the top-level declaration. The Java class implementing a declaration is found at run time using Java's normal classpath-based lookup mechanism.

7.4 Applications

An application is any globally-visible function that takes no arguments.
If the user types at the command line:

```
plaid Name
```

where `Name` is a qualified name, the `plaid` executable will search the classpath for a declaration of the named function and will try to execute it.

References

James Gosling, Bill Joy, Guy Steele, and Gilad Bracha. *The Java Language Specification, Third Edition*. Addison Wesley, 2005.

Karl Naden, Robert Bocchino, Jonathan Aldrich, and Kevin Bierhoff. A type system for borrowing permissions. In *Proceedings of Principles of Programming Languages (POPL)*, 2012.

Sven Stork, Paulo Marques, and Jonathan Aldrich. Concurrency by default: using permissions to express dataflow in stateful programs. In *Onward!*, 2009.

Sven Stork, Jonathan Aldrich, and Paulo Marques. micro-AEmimium Language Specification. Technical Report CMU-ISR-10-125R1, Carnegie Mellon University, 2010.

Joshua Sunshine, Karl Naden, Sven Stork, Jonathan Aldrich, and Éric Tanter. First-class state change in plaid. In *Proceedings of Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA)*, 2011.