

Contents

1	Introduction	3
2	Empirical Analysis of Use-Cases	5
2.1	Unsafe Rust	6
2.2	Mutability	6
2.3	Conclusions	8
3	Foundations	11
3.1	Rust	11
3.2	Refinement Types	12
4	The MiniCorten Language	13
4.1	Features	13
4.1.1	Mutable References	13
4.2	Path Sensitivity	14
4.3	Strong Type Updates	14
4.3.1	Modularity	14
4.4	Mutual Reference	16
4.5	Atomic Updates	16
4.6	Syntax	16
4.7	Semantics	17
5	The Refinement Type System	19
5.1	Definitions	19
5.2	Expression Typing: $\Gamma \vdash e : \tau \Rightarrow \Gamma'$	20
5.3	Try: Statements $\Gamma \vdash s \Rightarrow \Gamma'$	21
5.4	Function Declaration Type Checking	22
5.5	Sub-Typing Rules: $\Gamma \vdash \tau \preceq \tau'$	22
5.6	Sub-Context Rules: $\Gamma \preceq \Gamma'$	22
5.7	Soundness of the Type System	23
5.8	Extensions	24
5.8.1	Records / structs	24
5.8.2	Predicate Generics	24
6	Implementation	25

7	Evaluation	27
7.1	Maximum using Path Conditions	27
7.2	Rephrasing builtins in terms of Refinement Types	27
7.3	Complex Mutable References	27
7.4	Proof of the Gauss Summation Formula	27
8	Related Work	29
8.1	Comparison to Flux	30
9	Conclusion & Future Work	31

Software correctness is a central goal for software development. One way to improve correctness is using strong and descriptive types and verifying them with type systems. In particular Refinement Types demonstrated, that even with a verification system that is restricted to a decidable logic, intricate properties can be expressed and verified in a lightweight and gradual way. However existing approaches for adapting Refinement Types from functional to imperative languages proved hard without compromising on at least one of core features of Refinement Types. This thesis aims to design a Refinement Type system without such compromises by taking advantage of Rust’s restriction to unique mutable references. I will define a Refinement Type language and typing rules and argue for their soundness. Additionally I will implement a prototype verifier to evaluate the effectiveness of the approach on selected examples.

Chapter 1

Introduction

With increasing amount and reliance of software, ensuring the correctness of programs is a vital concern for the future of software development. Although research in this area has made good progress, most approaches are not accessible enough for general adoption by the developers. Especially in light of a predicted shortage of developers[breaux_2021_2021-1], it is not sufficient to require developers to undergo year-long training in specialized and complex verification methods to ensure the correctness of their software. It is therefore crucial to integrate with their existing tooling and workflows to ensure the future high quality of software. One avenue for improving accessibility for functional verification is extending the expressiveness of the type system to cover more of the correctness properties. Using type systems for correctness was traditionally prevalent in purely functional languages where evolving states are often represented by evolving types, offering approachable and gradual adoption of verification methods. Tracing evolving states in the type system would be especially useful for languages with mutability, since substantial parts of the behaviour of the program is expressed as mutation of state. In particular Rust seems like a promising target language, because mutability is already tracked precisely and thus promising functional verification for relatively minimal effort on the programmer's part.

The goal of the thesis is to show that Refinement Types can be idiomatically adapted to languages with unique mutable references. The type system presented in this thesis enables gradual adoption of lightweight verification methods in mutable languages.

The type system is sound and effective in the identified use-cases. A feasibility study on minimal examples of challenging use cases shows how useful the proposed verification system is.

Specifying or verifying complete Rust modules or the entire Rust language is not the goal of the thesis. In particular `unsafe` Rust will not be taken into account in specification nor implementation. Implementing Liquid Type inference is also not a goal of the thesis.

The accompanying implementation extends the Rust compiler, enables auto-

matic parsing of the refinement type language and automated type checking of a subset of Rust, as well as limited inference and error reporting. The thesis also gives a description of the syntax and semantics of the subset of Rust as well as the refinement type system.

The contributions of this thesis are:

1. Automatic empirical analysis of the usage of mutability and unsafe in Rust using syntactic information
2. Extension to the Rust type system to allow for refinement type specifications
3. Description and implementation of a type checker for the introduced type system
4. Evaluation of the type system on minimal benchmarks

The thesis is structured as follows: In chapter XXX an empirical analysis of `unsafe` and mutability uses is performed. Then chapter XXX will give an overview of the foundation the thesis build upon. Chapter XXX defines the subset of Rust, that will be the basis for the type system. Next chapter XXX explains the actual type system and verification, as well as justifying its correctness, followed by chapter XXX, which will provide more information about the implementation. The type system will then be tested in minimal benchmarks in chapter XXX. Chapter XXX reports on related work and alternative approaches. Finally chapter XXX concludes the thesis and gives an overview over possible future work.

Chapter 2

Empirical Analysis of Use-Cases

Before designing a system for Rust, it makes sense to gain some understanding of how Rust is used. For this purpose we will look at two key features of Rust, that influence how an approachable verification should look like. Firstly `unsafe` Rust with similar guarantees to C would make verification and specification significantly harder. But if the use of `unsafe` is limited, like intended by the language designers, it would allow us to focus on the safe part of Rust and leave the verification of `unsafe` Rust to more complex verification systems. Secondly with mutability being the main difference to the traditional domain of Refinement Types, showing the need for covering this area and to what degree of approachability it is interesting.

There are two fundamental assumptions for the usefulness of the type system:

1. Uses of `unsafe` are rare
2. In the rare case, that `unsafe` is used, it is mostly used internally (not exposed to the callee)
3. Mutability is used pervasively in Rust.

To check these assumptions an analysis of existing Rust code was performed. As a basis for the analysis, the Rust package registry <https://www.crates.io> was used. It contains the source code for both Rust libraries as well as various applications written in Rust (e.g. `ripgrep`).

We analyzed all published Rust crates on February 2nd 2022 (Rust's version packages) on `crates.io` with at least 10 versions¹, which totals 11 882 crates, containing 228 263 files with a combined code-base size of over 64 million lines of Rust code (without comments and white space lines)².

¹The limit of 10 versions is used to eliminate inactive and placeholder packages

²Calculated with `cloc`

The analysis parses these files and searches for certain patterns, which are subsequently extracted and saved. Thanks to using the tree-sitter parsing framework, the analysis framework can easily be extended to other queries and languages.

2.1 Unsafe Rust

Firstly we will be checking the hypothesis 1. There is already some research on how `unsafe` is used in Rust. For example Astrauskas et al. [astrauskas_how_2020] found, that about 76% of crates did not use any unsafe. On top of that, unsafe signatures are only exposed by

”The majority of crates (76.4%) contain no unsafe features at all. Even in most crates that do contain unsafe blocks or functions, only a small fraction of the code is unsafe: for 92.3% of all crates, the unsafe statement ratio is at most 10%, i.e., up to 10% of the codebase consists of unsafe blocks and unsafe functions.” [astrauskas_how_2020] Our data seems to confirm this: 8 044 of the 11 882 crates (67.7%) did not use any unsafe.

Astrauskas et al. also found, that ”however, with 21.3% of crates containing some unsafe statements and – out of those crates – 24.6% having an unsafe statement ratio of at least 20%, we cannot claim that developers use unsafe Rust sparingly, i.e., they do not always follow the first principle of the Rust hypothesis.” [astrauskas_how_2020] If this is the case for applications, it might disprove hypothesis hypothesis 2. When checking unsafely for our use case, it makes sense to further distinguish between libraries and executables crates: Libraries are intended to be used by other Rust programs: Usage of unsafe in libraries may not be as problematic as in executables, because libraries are written once but used by many applications, justifying higher verification effort.

In our analysis we found, that firstly crates.io contains significantly more libraries than binaries³ (see Figure 2.1). And secondly libraries are much more likely to use `unsafe` Rust. Table shows the result of our analysis. Where uses of `unsafe` are counted and grouped by crate type (see Table 2.1). The data in the table includes all crates except `windows-0.32.0`, which alone contains 233 608 uses of `unsafe`. Nearly 2/3 of all other library `unsafe` combined. Looking at the distribution of unsafe uses in Figure 2.2, we can see, that this is an exception: Most other libraries do not use that much unsafe statements. We can also see, that even if a executable crate uses `unsafe`, it uses few.

2.2 Mutability

Finally we will check hypothesis 3, which asserts, that mutable variables and references are used pervasively in Rust. This was checked by analysing the

³Libraries and Executables are distinguished by checking if they contain `bin` or `lib` target or one of the corresponding files according to the cargo naming convention.

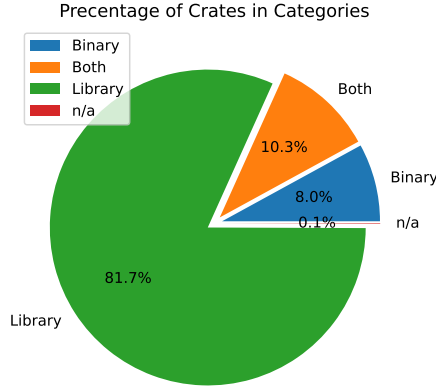


Figure 2.1: Percentage of Crates, that Contain Libraries, Executables or Both

Crate Type	Number of Unsafe Uses
Library	382 997
Both	7 720
Binary	930
n/a	215

Table 2.1: Number of Unsafe Uses by Crate Category

dataset for certain syntactical structures to infer mutability information about the following various AST items:

- **Local Variable Definitions** can be tracked with high confidence. They occur in function bodies and take the form: `let mut a = <expr>`
- **Parameters**, which are considered immutable if they are passed as immutable references or owned. They take the syntactic form: `mut a: i32`
- **Function Definitions**, which are considered immutable, if all parameters considered immutable.
- **Arguments** are parts of a function call and can be arbitrary expression, which makes the tracking hard.
- **Function Calls**, which are considered immutable, if all arguments are considered immutable.

A total of around 52 million of these items were found in the dataset.

Figure 2.3 shows the ratio of mutable to immutable items. For each syntactical category, the percentages are given relative to different objects:

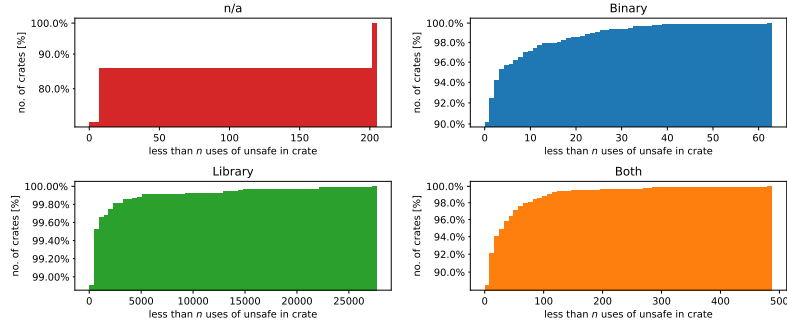


Figure 2.2: Cumulative, logarithmic histogram of the amount of `unsafe` uses in each category

1. Total: number of occurrences
2. Function: number of unique functions. I.e. 76.54% of functions have a immutable parameter.
3. File: number of unique files

Unfortunately not there are some areas, where the syntactic analysis is not sufficient. Namely the analysis of function calls and arguments, which is mainly caused by the uncertainty of mutability of complex arguments. Luckily the data from function definitions and parameters can complete the picture: About 80% of parameters are immutable and between about 10 and 20% of parameters are mutable. And less than 10% of functions have mutable parameters at all. When it comes to local variables, Rust users are more liberal in their use of mutability: About 30% of local variables are defined mutable.

For verification this means, that the use of mutability is wide spread. Especially local variables are often mutable and should therefore a verification system should try to minimize the effort for the user. Mutable parameters are less common, but still need to be accounted for in verification.

2.3 Conclusions

For this thesis the following conclusion can be drawn:

- Even though uses of `unsafe` are not rare, it is acceptable to ignore in favour of a simpler type system.
- Mutable variables are used very often and should therefore have very little associated specification effort.
- Mutable parameters are used less frequently justifying a higher specification effort, but their specification must still be possible.

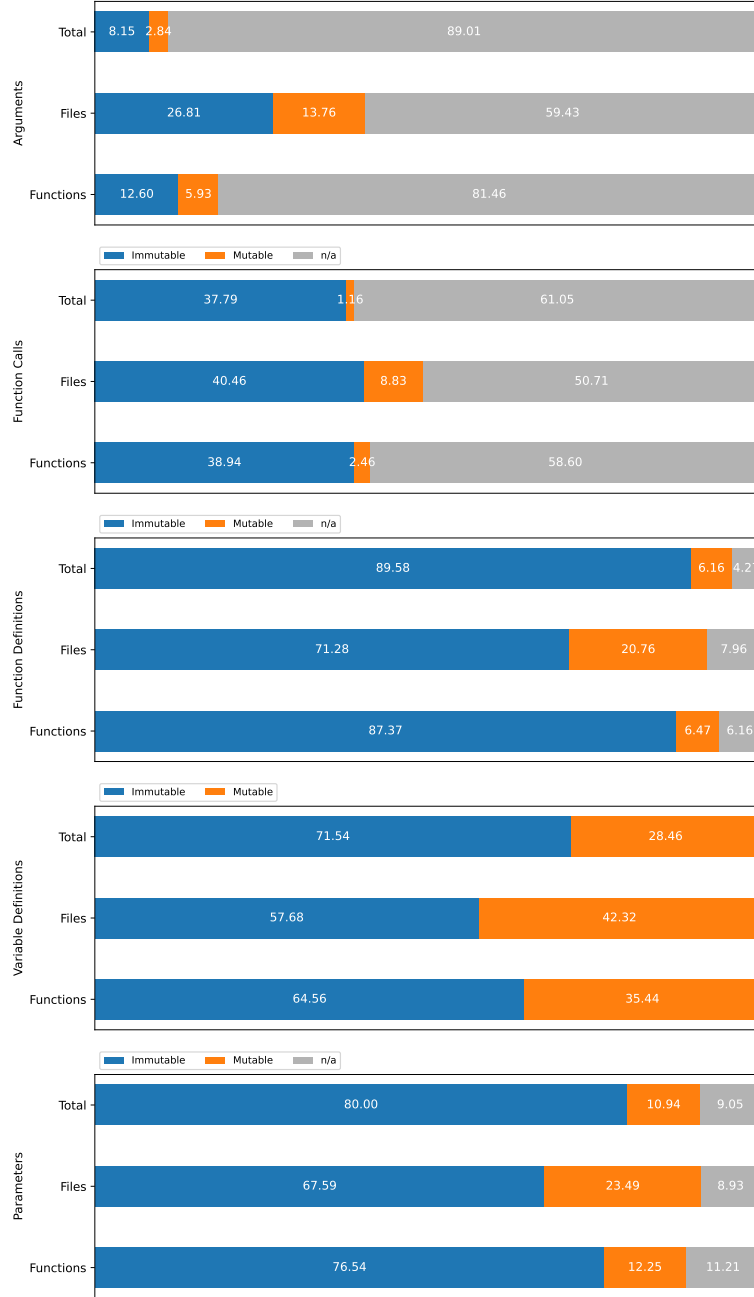


Figure 2.3: Ratio of Immutable to Mutable Versions of Different AST Items. Items are Counted by Unique Occurrences

Chapter 3

Foundations

3.1 Rust

Aliasing XOR Mutability The key behind Rust's type system is, that for every variable at every point during the execution, that variable is either aliased or mutable, but never both at the same time.

Rust accomplished this by introducing three types of access permissions for a variable, which are associated with every scope¹:

1. The scope *owns* the variable. This guarantees, that no other scope has access to any part of the variable and allows the scope to create references to (part of) the variable.
2. The scope (immutably) *borrow*s the variable. This guarantees, that as long as the variable is used, its *value will not change* and allows the scope to further lend the variable to other scopes.
3. The scope *mutably borrow*s the variable. This guarantees, that there are no other active references to any part of the memory of the variable.

A consequence of these rules is, that at any time, every piece of memory has a unique and compile-time known owner. There are no reference cycles.

This makes both aliasing as well as mutability quite tame: If a variable is aliased, it must be immutable and as a result, represents just a value (like in pure functional languages). If a variable is mutable, it can not be aliased and as a result, any effect of the mutation is well known and locally visible.

The Case for Rust as a Target Language Rust justifies a new approach to mutable verification, because it is possible to take advantage of guarantees provided by (safe) Rust's ownership system.

Rust is split into two languages: safe and unsafe Rust. Like most type systems, Rust's type and ownership system is conservative, meaning any (safe)

¹In modern Rust, this model was generalized to cover more cases, but the idea is still valid

Rust program that type checks, will not crash due to memory safety or type errors. Unsafe Rust gives the programmer the ability to expand the programs accepted by Rust outside that known-safe subset.

In this thesis, we will only consider safe Rust, which is the subset of Rust most programmers interact with (see section 2).

Rust features a few unusual design decisions that profoundly influence the design of verification systems for it. The following paragraphs will elaborate on this.

Opaque Generics Rust does not provide a way to check a generic parameter for its instantiation. This means that we cannot extract any additional information from a generic parameter T . A function from T to T can therefore only be the identity function. Wadler [wadler_theorems_1989] shows, that it is possible to derive facts about the behaviour of such polymorphic functions.

In contrast, languages with instance-of-checks allow an implementation of a polymorphic function to distinguish between different instances of the generic parameter, precluding this extensive reasoning.

Explicit Mutable Access A consequence of the ownership rules is, that a function can only mutate (part of) the state, that is passed to it as a parameter. There is no global state and there is no implicit access to an object instance. Thus any intention of mutating state must already be expressed in the function signature, which makes the specification of this mutation quite natural.

What Rust does not solve Even though Rust simplifies reasoning about mutability and aliasing of mutable data, Rust does not eliminate the need to reason about multiple references to mutable data. For example, Listing 3.1 shows how `cell`'s type is influenced by changes to `r`. This does not mean, that the ownership rule *mutability XOR aliasing* are broken: Even though both `cell` and `r` are mutable, but only one is active.

```
let mut cell = 2;
let r = &mut cell;
r += 1;
assert (cell , 2)
```

3.2 Refinement Types

Chapter 4

The MiniCorten Language

Rust’s main disadvantage as a target language is its size: There is a lot of syntax and semantics that would need to be accounted for. A lot of it is even incidental to the verification. To reduce the complexity and amount of work, that needs to be done, we will focus on a subset of Rust described in this section.

The goal is to remove as much incidental complexity as possible without compromising the central topic of research: How to extend LiquidTypes to mutability under the presence of Rust’s ownership model.

4.1 Features

This subsection will explain some of the key features of the Corten type system. Corten is directly embedded in Rust using two macros. Firstly the macro `ty!` can be used in place of a Rust type and adds a predicate to the Rust type, that any inhabitant of that type must satisfy. For example, the type `ty!{ v : i32 | v >= 0 }` stipulates, that a value of Rust type `i32` is positive. The second macro is `relax_ctx!{ ... }` which will be explained in subsection 4.4.

4.1.1 Mutable References

The key difference to Refinement Type Systems for functional languages to Rust is the pervasive use of mutation. Therefore the most important feature of Corten is the support for mutation and also mutable references.

The problem with references is the possibility of interdependencies: If one variable is changed, it might effect whether another variable is typed correctly. In listing 1 the return value `a` is effected by changes done to a different variable `a`. Conservative approximation requires, that all possible effects must be tracked.

Corten can accurately track references by phrasing them as predicates. For example the expression `&mut a` will be typed as `ty!{ _1 : &mut b | _1 == &a }` (where `b` is the Rust type of `a`). This is sensible, because in Rust’s ownership system, `a` must be the unique owner of the memory belonging to `a`. A

more involved example is given in the evaluation 7.2

Corten has two ways of dealing with assignment to mutable references: If the reference destination is known, the destination's type will be updated with the assigned values type (i.e. strong update). If there are multiple possible reference destinations, Corten will require the assigned value to satisfy the predicates of all possible destinations (i.e. weak updates). This is a standard approach (For example [kloos_asynchronous_2015]), but more precise in Rust: The set of possible destinations only grows, when it depends on the execution of an optional control flow path. This means most of the time, strong updates are possible and weak updates are only needed if the reference destination is actually dynamic (i.e. dependent on the execution) ¹

- importance of tracing return of mutable references

```
fn client() -> ty!{ v: i32 | v == 4 } {
  let a = 2; let b = &mut a;
  *b = 0; // changes a's value
  let c = &mut b;
  **c = 4; // changes a's value
  a
}
```

Listing 1: Example demonstrating interdependencies between mutable references

4.2 Path Sensitivity

Firstly, the type system is path sensitive, meaning that the type system is aware of necessary XXX that need to be passed for an expression to be evaluated. For example listing 2 shows a function computing the maximum of its inputs. In the then branch, a is only a maximum of a and b , because the condition $a > b$ implies it. Corten will symbolically evaluate the condition and store it in its typing context.

- inference - also with mutations

4.3 Strong Type Updates

4.3.1 Modularity

For a verification system to be scalable, it needs to be able modularize a proof. Corten can propagate type information across function calls and by taking advantage of Rust's ownership system, it can do so very accurately. Listing 4

¹It would also be possible to encode the path condition in the reference type, but this was decided against for the stated goal of simplicity for the user

```

fn max(a : ty!{ av: i32 }, b: ty!{ bv: i32 })
  -> ty!{ v: i32 | v >= av && v >= bv } {
  if a > b {
    a as ty!{ x : i32 | x >= av && x >= bv }
  } else {
    b
  }
}

```

Listing 2: Function computing the maximum of its inputs; guaranteeing that the returned value is larger than its inputs

```

fn strong_updates(b: ty!{ bv: i32 | bv > 0 }) -> ty!{ v: i32 | v > 2 } {
  let mut a = 2;
  a = a + b;
  a
}

```

Listing 3: Example of changes to `a`'s value effecting its type

shows, how an incrementing function `inc` can be specified: `inc` signature stipulates, that it can be called with any `i32` (given the name `a1`) and will return a value `a2`, which equals `a1 + 1`. Only the signature of `inc` is necessary for the type checking of `client`.

Notice, that all of the type information about `y` is preserved when `inc(x)` is called. There are no further annotations needed to type check this program. This is possible, because in safe Rust, any (externally observable) mutation done by a function must be part of the function signature. Corten expands on this, by enabling the user specify exactly how a referenced value is mutated.

```

fn inc(a: &mut ty!{ a1: i32 => a2 | a2 == a1 + 1 }) {
  *a = *a + 1;
}

fn client(mut x: ty!{ xv: i32 | xv > 2 }) -> ty!{ v: i32 | v > 7 } {
  let mut y = 2;
  inc(&mut x); inc(&mut x);
  inc(&mut y)
  x + y
}

```

Listing 4: Example showing how Corten allows for accurate type checking in the presence of function calls

4.4 Mutual Reference

Complex mutation patterns often result in complex interdependencies. We deem it necessary to allow different types to refer to each other.

why?

```
let sum = 0;
let i = 0;
```

Listing 5: Function computing the maximum of its inputs; guaranteeing that the returned value is larger than its inputs

4.5 Atomic Updates

- relax ctx - non-atomic might not be sufficient

4.6 Syntax

Call it CortenRust (after Weathering Steel)

What does the abbrev stand for? XXX Normal Form

This subsection will introduce the syntax of MiniCorten, a language modelled after a simplified version of Rust, with the addition of refinement types. To simplify the formal definitions and proofs, the language is restricted to ANF, meaning arguments of expressions must be variables. Note that the implementation does not have this restriction.

<i>program</i>	<code>::= func_decl *</code>	<i>function declaration</i>
<i>func_decl</i>	<code>::= ident(param *) -> ty { stmt }</code>	
<i>param</i>	<code>::= ident : ty</code>	
<i>stmt</i>	<code>::= expr</code>	<i>expression</i>
	<code> let mut? ident = expr</code>	<i>declaration</i>
	<code> ident = expr</code>	<i>assignment</i>
	<code> while (expr) { stmt }</code>	<i>while loop</i>
	<code> relax_ctx!{ pred* ; (ident : ty)* }</code>	<i>context relaxation</i>
<i>expr</i>	<code>::= ident</code>	<i>variable reference</i>
	<code> lit</code>	<i>constant</i>
	<code> expr + expr</code>	<i>addition</i>
	<code> ident(ident *)</code>	<i>function call</i>
	<code> if expr { stmt } else { stmt }</code>	<i>if expression</i>
	<code> stmt; expr</code>	<i>sequence</i>
	<code> * ident</code>	<i>dereference</i>
	<code> & ident</code>	<i>immutable reference</i>
	<code> &mut ident</code>	<i>mutable reference</i>
	<code> expr as ty</code>	<i>type relaxation</i>
<i>ty</i>	<code>::= ty!{ logic_ident : base_ty pred }</code>	<i>refinement type</i>
<i>pred</i>	<code>::= ref_pred</code>	<i>predicate for a reference type</i>
	<code> value_pred</code>	<i>predicate for a value type</i>
<i>ref_pred</i>	<code>::= logic_ident = & ident</code>	
	<code> ref_pred ref_pred</code>	<i>mutable reference</i>
<i>value_pred</i>	<code>::= logic_ident</code>	<i>variable</i>
	<code> pred && pred</code>	<i>conjunction</i>
	<code> pred pred</code>	<i>disjunction</i>
	<code> ! pred</code>	<i>negation</i>
<i>base_ty</i>	<code>::= i32</code>	<i>integer</i>
	<code> unit</code>	<i>unit type</i>
	<code> bool</code>	<i>boolean</i>
	<code> & base_ty</code>	<i>immutable reference</i>
	<code> &mut base_ty</code>	<i>mutable reference</i>
<i>lit</i>	<code>::= 0,1,...,n</code>	<i>integer</i>
	<code> true</code>	<i>boolean true</i>
	<code> false</code>	<i>boolean false</i>
	<code> ()</code>	<i>unit value</i>

4.7 Semantics

The semantics are mostly standard. The main difference being that Rust and our simplified language enable most statements to be used in place of an expression. The value is determined by the last statement in the sequence. For example If-Expressions, sequences of statements and function all follow this rule: Their (return) value is determined by the last statement or expression in the sequence.

The semantics loosely based on Jung's MiniRust.

Another difference between Rust and MiniCorten is of course the addition of refinement types.

In terms of the formal description, the rules are similar to Pierce's [pierce_types_2002-3] Reference language. The main difference is, that in Rust, every piece of data has a unique, known owner. This fact makes the concept of locations redundant. Instead we treat **ref** x as a value itself. The following definitions show the new execution rules.

Definition 4.7.1 (Execution-State). The state of execution is given by $\sigma : \text{PVar} \rightarrow \text{Value}$. The set of function declarations is constant and given by $\Sigma : \text{Fn-Name} \rightarrow ((\text{Arg}_1, \dots, \text{Arg}_n), (\text{ReturnValue}))$

Definition 4.7.2 (Small-Step Semantics of MiniCorten: $\langle e \mid \sigma \rangle \rightsquigarrow \langle e' \mid \sigma' \rangle$).

$$\begin{array}{c}
\text{SS-ASSIGN} \frac{\star}{x = v \mid \sigma \rightsquigarrow \mathbf{unit} \mid \sigma[x \mapsto v]} \quad \text{SS-ASSIGN-INNER} \frac{t \mid \sigma \rightsquigarrow t' \mid \sigma'}{x = t \mid \sigma \rightsquigarrow x = t' \mid \sigma'} \\
\\
\text{SS-DECL} \frac{\star}{\langle \mathbf{let} \ x = v \mid \sigma \rangle \rightsquigarrow \langle \mathbf{unit} \mid \sigma[x \mapsto v] \rangle} \\
\\
\text{SS-DEREF} \frac{\sigma(x) = v}{*\mathbf{ref} \ x \mid \sigma \rightsquigarrow v \mid \sigma} \quad \text{SS-DEREF-INNER} \frac{t \mid \sigma \rightsquigarrow t' \mid \sigma'}{*t \mid \sigma \rightsquigarrow *t' \mid \sigma'} \\
\\
\text{SS-REF-INNER} \frac{t \mid \sigma \rightsquigarrow t' \mid \sigma'}{\mathbf{ref} \ t \mid \sigma \rightsquigarrow \mathbf{ref} \ t' \mid \sigma'} \\
\\
\text{SS-SEQ-INNER} \frac{e_1 \mid \sigma \rightsquigarrow e'_1 \mid \sigma'}{e_1; e_2 \mid \sigma \rightsquigarrow e'_1, e_2 \mid \sigma'} \quad \text{SS-SEQ-N} \frac{\star}{\mathbf{unit}; e_2 \mid \sigma \rightsquigarrow e_2 \mid \sigma'} \\
\\
\text{SS-IF-T} \frac{\sigma(x) = \mathbf{true}}{\langle \mathbf{if} \ x \{e_t\} \ \mathbf{else} \ \{e_e\} \mid \sigma \rangle \rightsquigarrow \langle e_t \mid \sigma \rangle} \quad \text{SS-IF-F} \frac{\sigma(x) = \mathbf{false}}{\langle \mathbf{if} \ x \{e_t\} \ \mathbf{else} \ \{e_e\} \mid \sigma \rangle \rightsquigarrow \langle e_e \mid \sigma \rangle} \\
\\
\text{SS-WHILE} \frac{\star}{\langle \mathbf{while} \ e_c \{e_b\} \mid \sigma \rangle \rightsquigarrow \langle \mathbf{if} \ e_c \{e_b; \mathbf{while} \ e_c \{e_b\}\} \ \mathbf{else} \ \{\mathbf{unit}\} \mid \sigma \rangle}
\end{array}$$

Chapter 5

The Refinement Type System

5.1 Definitions

P is the set of program variables used in rust program. Common names a, b, c

L is the set of logic variables used in refinement types. Common names: α, β

$\Gamma = (\mu, \Phi)$ is a tuple containing a function $\mu : P \rightarrow L$ mapping all program variables to their (current) logic variable and a set of formulas Φ over L . During execution of statements, the set increases monotonically

τ is a user defined type $\{\alpha : b \mid \varphi\}$. Where α is a logic variable from L , b is a base type from Rust (like `i32`) and φ is a formula over variables in L .

Abbreviations We write:

- Γ, c for $(\mu, \Phi \wedge c)$
- $\Gamma[a \mapsto \alpha]$ for $(\mu[a \mapsto \alpha], \Phi)$

5.2 Expression Typing: $\Gamma \vdash e : \tau \Rightarrow \Gamma'$

$$\begin{array}{c}
\text{LIT} \frac{l \text{ fresh} \quad \text{base_ty}(v) = b}{\Gamma \vdash v : \{l : b \mid l \doteq v\} \Rightarrow \Gamma} \\
\text{VAR} \frac{\alpha \text{ fresh} \quad \mu(x) = \beta}{\Gamma = (\mu, \Phi) \vdash x : \{\alpha : b \mid \beta \doteq \alpha\} \Rightarrow \Gamma} \\
\text{VAR-REF} \frac{\Gamma \vdash y : \tau \quad \Gamma \vdash x : \{\beta : \&b \mid \beta \doteq \&y\}}{\Gamma \vdash *x : \tau \Rightarrow \Gamma} \\
\text{IF} \frac{\Gamma \vdash \mu(x) : \text{bool} \Rightarrow \Gamma \quad \Gamma, \mu(x) \doteq \mathbf{true} \vdash e_t : \tau \Rightarrow \Gamma' \quad \Gamma, \mu(x) \doteq \mathbf{false} \vdash e_e : \tau \Rightarrow \Gamma'}{\Gamma \vdash \text{if } x \text{ then } e_t \text{ else } e_e : \tau \Rightarrow \Gamma'} \\
\text{WHILE} \frac{\Gamma_I, \mu(x) \doteq \mathbf{true} \vdash s \Rightarrow \Gamma'_I \quad \Gamma'_I, \mu(x) \doteq \mathbf{true} \preceq \Gamma_I}{\Gamma_I \vdash \mathbf{while } x \{s\} \Rightarrow \Gamma_I, \mu(x) \doteq \mathbf{false}} \\
\text{SEQ} \frac{\Gamma \vdash s_1 : \tau_1 \Rightarrow \Gamma' \quad \Gamma' \vdash \bar{s} : \tau \Rightarrow \Gamma''}{\Gamma \vdash s_1; \bar{s} : \tau \Rightarrow \Gamma''} \\
\text{ADD} \frac{\gamma \text{ fresh} \quad \mu(x_1) = \alpha \quad \mu(x_2) = \beta}{\Gamma \vdash x_1 + x_2 : \{\gamma : b \mid \gamma \doteq \alpha + \beta\} \Rightarrow \Gamma} \\
\text{DECL} \frac{\Gamma \vdash e : \{\beta \mid \varphi\} \Rightarrow \Gamma'}{\Gamma \vdash \mathbf{let } x = e : () \Rightarrow \Gamma'[x \mapsto \beta], \varphi} \\
\text{ASSIGN} \frac{\Gamma \vdash e : \{\beta \mid \varphi\} \Rightarrow \Gamma'}{\Gamma \vdash x = e : \mathbf{unit} \Rightarrow \Gamma'[x \mapsto \beta], \varphi} \\
\text{ASSIGN-STRONG} \frac{\Gamma \vdash e : \{\beta \mid \varphi\} \Rightarrow \Gamma' \quad \Gamma \vdash x : \{\alpha : \&b \mid \alpha \doteq \&y\}}{\Gamma \vdash *x = e : \tau \Rightarrow \Gamma'[y \mapsto \beta], \varphi} \\
\text{ASSIGN-WEAK} \frac{\Gamma \vdash e : \tau \Rightarrow \Gamma' \quad \Gamma' \vdash x : \{\alpha : \&b \mid \alpha \doteq \&y_1 \vee \dots \vee \alpha \doteq \&y_n\} \Rightarrow \Gamma' \quad \Gamma' \vdash \tau \preceq \{\beta_1 \mid \varphi_1\} \quad \Gamma'[x \mapsto \beta_1], \varphi_1 \preceq \Gamma' \quad \dots}{\Gamma \vdash *x = e : \tau \Rightarrow \Gamma'} \\
\text{FN-CALL} \frac{\text{subst} = \dots \quad (\mu[\text{subst}], \alpha \doteq \mu(a) \wedge \dots \wedge \varphi_\alpha \wedge \dots) \preceq (\mu, \Phi) \quad f : (\{\alpha \mid \varphi_\alpha\} \Rightarrow \{\alpha' \mid \varphi'_\alpha\}, \dots) \rightarrow \tau}{(\mu, \Phi) \vdash f(a, \dots, \&\text{mut}b, \dots) : \tau \Rightarrow (\mu[a \mapsto \alpha', \dots, \text{subst}^{-1}], \Phi \wedge \varphi'_\alpha \wedge \dots)} \\
\text{INTRO-SUB} \frac{\Gamma \vdash e : \tau \quad \Gamma \vdash \tau \preceq \tau'}{\Gamma \vdash e \text{ as } \tau' : \tau'}
\end{array}$$

5.3 Try: Statements $\Gamma \vdash s \Rightarrow \Gamma'$

$$\begin{array}{c}
\text{IF} \frac{\Gamma \vdash \mu(x) : \text{bool} \Rightarrow \Gamma \quad \Gamma, \mu(x) \doteq \text{true} \vdash s_t \Rightarrow \Gamma' \quad \Gamma, \mu(x) \doteq \text{false} \vdash s_e \Rightarrow \Gamma'}{\Gamma \vdash \text{if } x \text{ then } s_t \text{ else } s_e \Rightarrow \Gamma'} \\
\\
\text{WHILE} \frac{\Gamma_I, \mu(x) \doteq \text{true} \vdash s \Rightarrow \Gamma'_I \quad \Gamma'_I, \mu(x) \doteq \text{true} \preceq \Gamma_I}{\Gamma_I \vdash \text{while } x \{s\} \Rightarrow \Gamma_I, \mu(x) \doteq \text{false}} \\
\\
\text{SEQ} \frac{\Gamma \vdash s_1 \Rightarrow \Gamma' \quad \Gamma' \vdash s_2 \Rightarrow \Gamma''}{\Gamma \vdash s_1; s_2 \Rightarrow \Gamma''} \\
\\
\text{DECL} \frac{\Gamma \vdash e : \{\beta \mid \varphi\}}{\Gamma \vdash \text{let } x = e \Rightarrow \Gamma'[x \mapsto \beta], \varphi} \\
\\
\text{ASSIGN} \frac{\Gamma \vdash e : \{\beta \mid \varphi\}}{\Gamma \vdash x = e \Rightarrow \Gamma'[x \mapsto \beta], \varphi} \\
\\
\text{ASSIGN-STRONG} \frac{\Gamma \vdash e : \{\beta \mid \varphi\} \quad \Gamma \vdash x : \{\alpha : \&b \mid \alpha \doteq \&y\}}{\Gamma \vdash *x = e \Rightarrow \Gamma[y \mapsto \beta], \varphi} \\
\\
\begin{array}{c}
\Gamma \vdash e : \tau \quad \Gamma' \vdash x : \{\alpha : \&b \mid \alpha \doteq \&y_1 \vee \dots \vee \alpha \doteq \&y_n\} \\
\Gamma \vdash \tau \preceq \{\beta_1 \mid \varphi_1\} \quad \Gamma[x \mapsto \beta_1], \varphi_1 \preceq \Gamma \\
\vdots \\
\Gamma \vdash \tau \preceq \{\beta_n \mid \varphi_n\} \quad \Gamma[x \mapsto \beta_n], \varphi_n \preceq \Gamma'
\end{array} \\
\text{ASSIGN-WEAK} \frac{\Gamma \vdash \tau \preceq \{\beta_n \mid \varphi_n\} \quad \Gamma[x \mapsto \beta_n], \varphi_n \preceq \Gamma'}{\Gamma \vdash *x = e \Rightarrow \Gamma'} \\
\\
\text{FN-CALL} \frac{\text{subst} = \dots \quad (\mu[\text{subst}], \alpha \doteq \mu(a) \wedge \dots \wedge \varphi_\alpha \wedge \dots) \preceq (\mu, \Phi) \quad f : (\{\alpha \mid \varphi_\alpha\} \Rightarrow \{\alpha' \mid \varphi'_\alpha\}, \dots) \rightarrow \tau}{(\mu, \Phi) \vdash \text{let res} = f(a, \dots, \&\text{mutb}, \dots) \Rightarrow (\mu[a \mapsto \alpha', \dots, \text{subst}^{-1}], \Phi \wedge \varphi'_\alpha \wedge \dots)}
\end{array}$$

Expressions: $\Gamma \vdash e : \tau$

$$\begin{array}{c}
\text{LIT} \frac{l \text{ fresh} \quad \text{base_ty}(v) = b}{\Gamma \vdash v : \{l : b \mid l \doteq v\}} \\
\\
\text{VAR} \frac{\alpha \text{ fresh} \quad \mu(x) = \beta}{\Gamma = (\mu, \Phi) \vdash x : \{\alpha : b \mid \beta \doteq \alpha\}} \\
\\
\text{VAR-REF} \frac{\Gamma \vdash y : \tau \quad \Gamma \vdash x : \{\beta : \&b \mid \beta \doteq \&y\}}{\Gamma \vdash *x : \tau} \\
\\
\text{ADD} \frac{\gamma \text{ fresh} \quad \mu(x_1) = \alpha \quad \mu(x_2) = \beta}{\Gamma \vdash x_1 + x_2 : \{\gamma : b \mid \gamma \doteq \alpha + \beta\}} \\
\\
\text{INTRO-SUB} \frac{\Gamma \vdash e : \tau \quad \Gamma \vdash \tau \preceq \tau'}{\Gamma \vdash e \text{ as } \tau' : \tau'}
\end{array}$$

5.4 Function Declaration Type Checking

Without loss of generality, we assume arguments are ordered by their type: First immutable / owner arguments and then mutable arguments.

When handling mutable references in parameters some subtleties need to be considered. The function can change both the referenced *value* as well as the reference *location*. To describe the referenced value is normally done using the $\{\alpha \mid \alpha \doteq \&b\}$ syntax. The question arises: If α was the logic variable for a parameter, what should be the analog for b ? In contrast to local variables, there is no variable representing the referenced value for parameters. As seen in section 4.1, using the dereference operator would come with a lot of complications. Instead we introduce arg_n^i , a special variable denoting the initial abstract value (i.e. stack location), that the mutable reference of argument n points to. i denotes the level of nesting: For the n th parameter with the Rust type `&mut i32` we would generate arg_n^1 and arg_n^2 .

$$\text{FN-DECL} \frac{\begin{array}{c} (\{a_1 \mapsto \alpha_1, b_1 \mapsto \delta_1\}, \varphi_1^\alpha \wedge \dots \wedge \delta_1 = arg_1^1 \wedge \varphi_1^\beta \wedge \dots) \vdash \bar{s} \Rightarrow \Gamma' \\ \Gamma' \vdash s_{res} : \tau_{res} \Rightarrow \Gamma'' \quad \Gamma'' \vdash \tau_{res} \preceq \tau \quad \Gamma'' \preceq (\{\beta_1 \mapsto \gamma_1\}, \varphi_1^\gamma) \end{array}}{\text{fn } f(a_1 : \{\alpha_1 \mid \varphi_1^\alpha\}, \dots, b_1 : \{\beta_1 \mid \varphi_1^\beta \Rightarrow \gamma_1 \mid \varphi_1^\gamma\}) \rightarrow \tau\{\bar{s}; s_{res}\}}$$

5.5 Sub-Typing Rules: $\Gamma \vdash \tau \preceq \tau'$

$$\preceq\text{-TY} \frac{\Phi \wedge \varphi'[\beta \triangleright \alpha] \models \varphi}{\Gamma = (\mu, \Phi) \vdash \{\alpha \mid \varphi\} \preceq \{\beta \mid \varphi'\}}$$

alternative (should be equivalent):

$$\preceq\text{-TY-ALT} \frac{\Gamma[f \mapsto \alpha], \varphi \preceq \Gamma[f \mapsto \beta], \varphi' \quad f \text{ fresh}}{\Gamma \vdash \{\alpha \mid \varphi\} \preceq \{\beta \mid \varphi'\}}$$

5.6 Sub-Context Rules: $\Gamma \preceq \Gamma'$

$$\preceq\text{-CTX} \frac{\Phi'[\mu(\alpha) \triangleright \mu'(\alpha) \mid \alpha \in \text{dom}(\mu)] \models \Phi \quad \text{dom}(\mu) \subseteq \text{dom}(\mu')}{(\Phi, \mu) \preceq (\Phi', \mu')}$$

In contrast to other refinement type systems, Corten allows types in the context to refer to one another. For example a type specifications $\mathbf{a} : \{ \alpha : \text{i32} \mid \alpha > \beta \}$, $\mathbf{b} : \{ \beta : \text{i32} \mid \beta \neq \alpha \}$ would be valid and result in the context $\Gamma = (\emptyset, \alpha \neq \beta \wedge \beta \geq \alpha)$. In the example 0 is a subtype of \mathbf{a} 's type as long as $\mathbf{b} \geq 1$.

This means that types may need to be changed atomically.

5.7 Soundness of the Type System

As usual, we consider the three properties (see Pierce [pierce_types_2002]) of a type system when assessing the correctness of MiniCorten:

Definition 5.7.1. State Conformance: A state σ is conformant with respect to a typing context $\Gamma = (\mu, \varphi)$ (written as $\sigma : \Gamma$), iff:

$$\models \left(\bigwedge_{x \in \text{Var}} \mu(x) \doteq \sigma(x) \right) \rightarrow \varphi$$

I.e. the a execution's value assignments imply the type refinements

1. Progress: If t is closed and well-typed, then t is a value or $t \rightsquigarrow t'$, where t' is a value.
2. Preservation: If $\Gamma \vdash e : \tau \Rightarrow \Gamma$ and $e \rightsquigarrow e'$, then $\Gamma \vdash e' : \tau \Rightarrow \Gamma$
3. Preservation of State Conformance

need to show for program or term?

The state conformance rule differs from the usual rule in two ways. Firstly we consider the runtime value instead of the runtime type. Secondly there needs to be one set of variable assignments that satisfies all predicates in Γ .

Because MiniCorten is based on Rust, we will assume, that progress, as well as preservation for the base-types. This includes preservation of type- and ownership-safety. Thus it is sufficient to prove, that assuming these properties hold, preservation of the refinement types is holds.

Lemma 5.7.2. *Preservation of State Conformance: If $\Gamma \vdash e : \tau \Rightarrow \Gamma'$, $\sigma : \Gamma$ and $\langle t \mid \sigma \rangle \rightsquigarrow \langle t' \mid \sigma' \rangle$, then $\sigma' : \Gamma'$*

Proof. Rule Induction over $\langle t \mid \sigma \rangle \rightsquigarrow \langle t' \mid \sigma' \rangle$

- CASE SS-ASSIGN $\sigma' = \sigma[x \mapsto v]$. With rule inversion ASSIGN: **unit** and post state: $\Gamma'[x \mapsto \beta], \varphi$ with $\Gamma \vdash e : \{\beta \mid \varphi\}$.
 e is a value, because of the requirement in SS-ASSIGN. By rule inversion LIT over assumption $\Gamma \vdash v : \{\beta \mid \varphi\} \Rightarrow \Gamma$: $\varphi = (\beta \doteq v)$
 $\mu'(x) = \beta$, $\sigma'(x) = v$ and induction hypothesis $\sigma : \Gamma$. Since $\mu'(x) \doteq \varphi(x) \rightarrow \varphi$ is trivially valid. XXXBanane gives us the goal.
- CASE VAR

□

Lemma 5.7.3. *State Conformance is XXXBanane: If $\sigma : \Gamma$, $(\beta = v) \rightarrow \varphi$ then $\sigma[x \mapsto v] : \Gamma[x \mapsto \beta] \wedge \varphi$*

Proof.

$$\models \left(\bigwedge_{x \in \text{Var}} \mu(x) \doteq \sigma(x) \right) \rightarrow \varphi$$

...?

□

5.8 Extensions

5.8.1 Records / **structs**

5.8.2 Predicate Generics

Chapter 6

Implementation

- Type Aliases - working without special compiler - compiler plugin - IDE - HIR
- Location data (spans) - Additional features - basic inference - no ANF

Chapter 7

Evaluation

7.1 Maximum using Path Conditions

7.2 Rephrasing builtins in terms of Refinement Types

`panic`

`assert`

7.3 Complex Mutable References

7.4 Proof of the Gauss Summation Formula

Chapter 8

Related Work

There currently does not exist an implementation of refinement types for Rust.

Relevant papers originate from two lines of work. Firstly additions to refinement types for mutability, asynchronous execution etc. and secondly other verification frameworks for Rust.

For example, Lanzinger [lanzinger_property_2021] successfully adapted refinement types to Java, which allows the user to check, that property types described by java annotations hold true throughout the program. At this point in time, specification and verification is limited to immutable (`final`) data.

Kloos et al. [kloos_asynchronous_2015] extended refinement types to mutable and asynchronous programs. The paper explores how changes to possibly aliased memory cells can be tracked throughout a OCaml program. For that purpose the types are extended by a set of requirements on memory objects, which track distinctness and refined types of these memory cells. In contrast to OCaml, Rust already guarantees that mutable memory is not aliased and in particular *all mutable memory locations must be accessible by a variable name in the current context*, which offers substantial advantages in terms of simplicity to specification and verification of Rust programs.

Refinement types are also used in other applications. For example Graf et al. [graf_lower_2020] use refinement types to check the exhaustiveness of pattern matching rules over complex (G)ADT types in Haskell. To check the exhaustiveness of patterns in Liquid Rust with ADTs may require similar approaches.

In terms of alternative verification approaches, Prusti[astrauskas_leveraging_2019] is notable, because of their work on formalizing the full Rust semantics, including `unsafe`. Prusti is a heavy-weight functional verification framework for Rust; based on separation logic.

Alternative verification approaches also exists: For example RustHorn[matsushita_rusthorn_2020] employs constrained horn clauses based verification to Rust. Particularity relevant for this thesis is the novel formalization for mutable references used in the paper. The authors stipulate that mutable references should be specified by a pre- and post-state from before a reference is borrowed to after it is returned.

The notion of Abstract Refinement Types that this thesis is based on is defined by Vazou et al. [**vazou_abstract_2013**]. The basic idea is to allow the programmer to "refine" language types with predicates from a decidable logic. The type system has a notion of subtyping for refined types, where one type is a subtype of another if one predicate implies the other.

8.1 Comparison to Flux

Chapter 9

Conclusion & Future Work