# Formalizing Model Inference of MicroPython*

Carlos Mão de Ferro
*LASIGE, Faculdade de Ciências,*
*Universidade de Lisboa*
Lisboa, Portugal

Tiago Cogumbreiro
*University of Massachusetts Boston*
Boston, MA, USA

Francisco Martins
*Universidade dos Açores*
Ponta Delgada, Portugal

August 10, 2023

**Abstract**

Model checking has often been used for verifying Cyber-Physical Systems (CPS). A major challenge is how to capture a model that represents the actual behavior of the software. Model extraction can introduce errors that can affect the accuracy of the analysis including loss of precision, inconsistency, non-conformance, and over- and under-approximations.

In this paper, we formalize and prove the correctness of extracting a model from a subset of the MicroPython programming language with respect to a trace-based semantics. The extracted models capture the order of method calls and can be model checked using Shelley. We formalize the extraction process from an intermediate representation of MicroPython codes and prove that the behavior of our intermediate representation is a regular language. Our formalization and theoretical results are fully mechanized using the Coq proof assistant.

*Edit by Tiago Cogumbreiro (January 2, 2024): added syntax of sequence to Figure 4; described syntax of sequence in §3.2. Paper originally published at the 1st International Workshop on Verification & Validation of Dependable Cyber-Physical Systems. DOI: 10.1109/DSN-W58399.2023.00069

# 1 Introduction

*Model extraction* [32, 45] is the process of automatically generating a model that captures the behavior of a system and its key properties. Model extraction is a balancing act between two approaches: *over-approximation* where the extracted model is too general and includes more behaviors than in the original system, *e.g.,* may trigger alarms that do not appear in practice; *under-approximation* where the extracted model is incomplete and misses certain behaviors, *e.g.,* may miss certain alarms. Further, the extraction process must carefully select a level of detail of a system: adding too much detail makes the model too big and impossible to analyze due to the state-explosion problem. Additionally, inaccuracies and bugs in the extraction process may lead to spurious behaviors that lead to incomplete or incorrect verification results [10].

Model extraction is particularly important in the context of cyber-physical system (CPS) applications, where the correctness and safety of the system are critical [28, 34, 38, 48, 42, 39, 47]. CPS applications typically involve the integration of physical systems, such as sensors, actuators, and control systems. This paper focuses on *formalizing* the model extraction process of CPS applications written in MicroPython [26] to be used in the context of model checking.

Formalizing model extraction has the following benefits. Firstly, the formalization offers a clear and precise description of the extraction process, which improves the understanding of the capabilities of our model checker. Secondly, having the extraction process formalized allows us to reason about properties of the source system and of the target model. Importantly, we can characterize the *expressiveness* of the extraction.

This paper specifies the model extraction of Shelley [17], a model-checking framework that features linear temporal logic on finite traces ($\text{LTL}_\text{f}$) [2, 19]. Similar tools exist for other languages including Java [27, 30, 9], C++ [6], and Lustre [24, 30] although their focus is usually on concurrency. In contrast, our temporal claims are written in terms of function calls that manipulate physical resources that can be specified with ordering constraints. MicroPython classes that directly manipulate physical resources (called constrained classes/objects) are annotated with ordering constraints, akin to typestates [51]. Further, and without resorting to any more ordering annotations, our analysis verifies the correct usage of constrained objects, and infers the ordering constraints of methods using other constrained objects. Unlike most model checkers, Shelley models do not capture state change of program-variables, nor the communication of concurrent systems [31, 46, 37, 24, 13, 7, 12].

In summary, our paper makes the following contributions:

1. We formalize the process of extracting a Shelley model from a small imperative language.

2. We prove the *correctness* of the extraction process: a trace is produced by the source program if, and only if, a trace is produced by the behavior of a program. We show that the behavior of a program is a regular language.

Table 1: Shelley's annotations, where to apply them in a MicroPython program, and their meanings.

| Annotation | Applies to | Meaning |
|---|---|---|
| `@claim` | class | temporal requirement |
| `@sys` | class | base `class` |
| `@sys(["`$s_1$`",. . . ,"`$s_n$`"])` | class | composite `class` |
| `@op_initial` | method | invoke in first place |
| `@op_final` | method | invoke in last place |
| `@op_initial_final` | method | invoke in first and last places |
| `@op` | method | invoke in between an initial and final methods |

3. The formalization and the results of this paper are fully mechanized using the Coq proof assistant and are available in [16].

The rest of the paper is organized as follows. Section 2 overviews our approach through a series of examples. Section 3 presents our model inference, formalizes the behavior extraction of a program (which represents the code of a method declaration), and establishes our main result of correctness. Section 4 discusses related work. Finally, Section 5 concludes the paper and considers the next steps in our research.

# 2 Model checking with Shelley

In this section, we detail how to use the Shelley framework to model check a MicroPython program. Our tool provides a MicroPython Application Programming Interface (API) that enables the automatic verification of the order of method calls in an object hierarchy. We use annotations, according to Table 1, so that developers can provide more information about the expected behavior of methods and prevent an unintended order of actions from occurring. This way, we enforce the behavior of an object and how it can be used in such a way that we can model check temporal requirements. Shelley includes a visualization tool that automatically generates behavior diagrams based on the code annotations and based on the control flow of the code under analysis.

To make our analysis scalable, Shelley makes two important design decisions: a restricted programming model, and an over-approximated behavior. Regarding the former, Shelley does not support non-terminating (infinite) behaviors, ignores object aliasing, and only considers method invocation of fields. Regarding the latter, our approach is closer to type checking of a behavioral type, than to model check the full behavior of a program. The code of each method must be expressed as a regular expression representing any possible sequence of method calls. Shelley features sequencing, nondeterministic choice, and terminating loops (via the Kleene-star operator). Further, our tool disregards the program's internal state, *e.g.,* the arguments of method calls, the condition used

Listing 2.1: Class `Valve`

```
1  @sys
2  class Valve:
3    def __init__(self):
4      self.control = Pin(27, OUT)
5      self.clean = Pin(28, OUT)
6      self.status = Pin(29, IN)
7
8    @op_initial
9    def test(self):
10     if self.status.value():
11       return ["open"]
12     else:
13       return ["clean"]
14
15   @op
16   def open(self):
17     self.control.on()
18     return ["close"]
19
20   @op_final
21   def close(self):
22     self.control.off()
23     return ["test"]
24
25   @op_final
26   def clean(self):
27     self.clean.on()
28     return ["test"]
```

to branch, and the loop bounds.

We now give a brief guide on the annotations that Shelley offers to verify a MicroPython class. Our running example is based on an industrial use case in [18], a battery-operated wireless controller that switches water valves according to a scheduled irrigation plan.

## 2.1 Specifying a class behavior

Valves are electromechanical devices that can be programmed to control water flow. Class `Valve` uses general-purpose input-output pins to operate a physical valve, as seen in Listing 2.1. Our verification goal is to minimize the chance of clogging the physical valve, so users of the `Valve` must test the status of the valve before opening it. Additionally, to conserve battery, we want to verify that users of `Valve` always test the status of the valve before cleaning up any debris. The class annotation `@sys`, in Line 1, tells Shelley to verify class `Valve`. The method annotations `@op_initial`, `@op`, and `@op_final` tell Shelley which methods to consider in the verification. We list our annotations in Table 1. To meet our verification goals, Shelley ensures that any users of `Valve` (*e.g.*, Listing 2.2) respect the specified method-ordering.

We now explain how to declare the method-ordering shown in Figure 1. The annotation `op_initial` ensures that after creating an instance of `Valve` the only method that can be invoked is `test` (Lines 9 to 13) — multiple methods can be declared as initial. Each annotated method must `return` the set of

Listing 2.2: Class BadSector

```
@claim("(!a.open) W b.open"))
@sys(["a", "b"])
class BadSector:
  def __init__(self):
    self.a = Valve()
    self.b = Valve()

  @op_initial_final
  def open_a(self):
    match self.a.test():
      case ["open"]:
        self.a.open()
        return ["open_b"]
      case ["clean"]:
        self.a.clean()
        print("a failed")
        return []

  @op_final
  def open_b(self):
    match self.b.test():
      case ["open"]:
        self.b.open()
        self.a.close()
        self.b.close()
        return []
      case ["clean"]:
        self.b.clean()
        print("b failed")
        self.a.close()
        return []
```

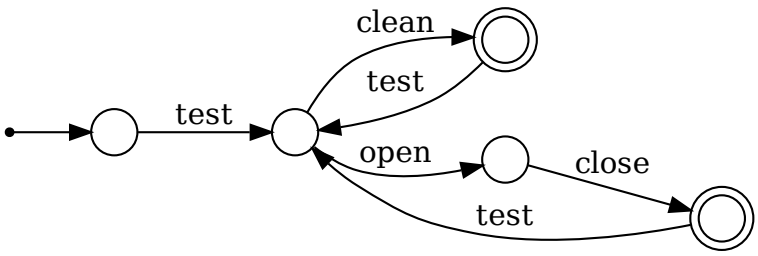


Figure 1: `Valve` diagram automatically generated by our tool based on the annotations of Listing 2.1.

Table 2: Examples of return statements and their meanings.

| Return statement | Meaning |
|---|---|
| `return ["close"]` | expecting method "close" to be invoked next |
| `return ["open", "clean"]` | expecting methods "open" or "clean" to be invoked next |
| `return ["close"], 2` | expecting operation "close" to be invoked next and return the integer value 2 |
| `return ["close"], True` | expecting method "close" to be invoked next and return the boolean value True |
| `return ["open", "clean"], 2` | expecting methods "open" or "clean" to be invoked next and return the integer value 2 |

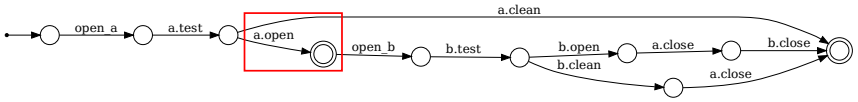


Figure 2: `BadSector` diagram (invalid usage of valves). After opening valve `a` we reach a possible final state, leaving valve `a` open and not respecting the `Valve` specification.

methods of `Valve` that can be invoked subsequently. For instance, after invoking `test` the return value can be `["open"]`, which means that the user must then invoke method `open`. Alternatively, after invoking `test` the return value can be `["clean"]`, so the user must invoke method `clean`. If a `return` statement allows several calls to follow it, we use the nomenclature `return ["`$m_1$`", ..., "`$m_n$`"]` and if no calls shall follow it we return an empty list. Finally, user-return values are also supported by using tuples, where the first position is reserved to specify the next available methods. Some examples are given in Table 2.

The decorator `op_final` allows for the declaration of "destructor" methods: method `close` must be the last method called, with respect to the object's lifetime. Importantly, since `open` is *not* marked as final, Shelley guarantees that the valve cannot be left open.

## 2.2 Verifying object usage

We now describe how Shelley verifies that class `BadSector` incorrectly uses instances of `Valve`. Further, we introduce the use of temporal claims to model

check the usage of verified instances. Shelley only considers the order of calls and disregards any values used in boolean conditions, loop ranges, and being passed in method calls. Shelley supports branching with `if/elif/else` and `match/case` and looping with `for` and `while`.

Class `BadSector` in Listing 2.2 handles the opening of valves in two separate methods — in the irrigation jargon, a *sector* is an irrigation zone where several water valves are grouped together. `BadSector` is a composite class since it uses two instances of `Valve`, `a` and `b`. Thus, depending on the method's annotations of `BadSector`, we might use the two valves in different ways. Shelley identifies the following invalid usage of `Valve`. Since method `open_a` is decorated with `initial` and `final`, this means that a user of `BadSector` could issue `open_a` without ever issuing `open_b`, potentially leaving valve `a` open, something that, according to the `Valve`'s specification, is an incorrect behavior, as depicted in Figure 2. Shelley outputs the following error message:

```
Error in specification: INVALID SUBSYSTEM USAGE
Counter example: open_a, a.test, a.open
Subsystems errors:
    * Valve 'a': test, >open< (not final)
```

**Matching exit points** In Lines 10 and 21 we intentionally use the `match` statement to distinguish the cases where the method being called has more than one exit point. For instance, as seen before, the `Valve`'s method `test` can be followed by either `open` or `clean`. Therefore, our tool checks if all possible exit points are being handled.

**Checking temporal requirements** Correctness claims express temporal properties on the code of the class being verified. Temporal claims are of great importance for software maintenance, as Shelley can check if changes to the class preserve the internal behavior being specified. Besides automatically verifying that each valve is being used according to the specification in Listing 2.1, Shelley also verifies temporal requirements. The temporal claim in Line 1 of Listing 2.2 states that valve `a` must remain closed until valve `b` is opened. The implemented behavior violates the temporal claim, so Shelley outputs the following error message:

```
Error in specification: FAIL TO MEET REQUIREMENT
Formula: (!a.open) W b.open
Counter example: a.test, a.open, b.test, b.open, a.close, b.close
```

The formula $\phi_1 \mathsf{W} \phi_2 = (\phi_1 \mathsf{U} \phi_2) \vee \mathsf{G}\, \phi_1$ is interpreted as a *weak until*, meaning that $\phi_1$ has to hold at least until $\phi_2$ or $\phi_1$ must remain true forever.

# 3 Model inference

In this section we detail the process of extracting a Shelley model from a MicroPython class. The model checking of a Shelley model is outside of the scope of this paper, so here we only focus on defining what constitutes a Shelley model. The model extraction process consists of the following steps:

1. **method dependency extraction:** (Section 3.1) our tool captures the

Listing 3.1: Class `Sector` with code elided to only show returns per method.

```
class Sector:
  def open_a(self):
    # ...
    return ["close_a", "open_b"]
    # ...
    return ["clean_a"]

  def clean_a(self):
    return ["open_a"]

  def close_a(self):
    # ...
    return ["open_a"]

  def open_b(self):
    # ...
    return []
    # ...
    return []
```

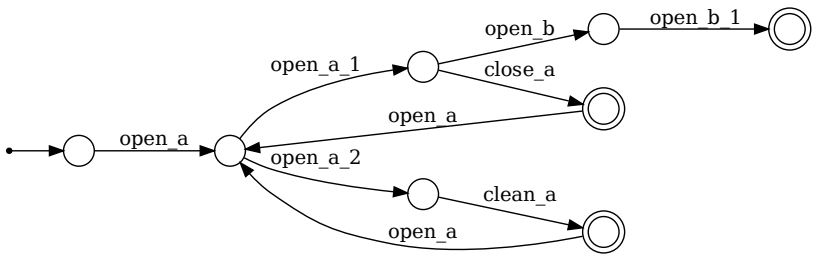


Figure 3: Shelley model of class `Sector`.

dependencies between methods being defined.

2. **method behavior extraction:** (Section 3.2) our tool extracts the behavior of each method being defined.

3. **method invocation analysis:** we check if method invocation of objects under analysis are defined in their respective classes; we also check for exhaustive tests on matches that take the result of a method-invocation under analysis.

## 3.1 Method dependency extraction

The method-dependency is defined as a directed graph, where the nodes represent the entry point of each method name and every exit point of a method; the arcs are ordering constraints. There is a single entry node per method. For instance, in Listing 3.1, we have 4 methods (`open_a`, `clean_a`, `close_a`, and `open_b`), so there are 4 entry nodes. In addition, there is an exit node per return in each method. For example, in Listing 3.1, method `open_a` has 2 return statements, thus we have 2 exit nodes: exit node, say (A), represents `return ["close_a", "open_b"]`, and the exit node, say (B), represents `return ["clean_a"]`. We link each entry node of a method to each of its exit nodes. For instance, in Figure 3, the entry node of `open_a` links to nodes (A) and (B). Finally, for each exit node and each method name being returned, we link the exit node to the entry node of the method being returned. That is, since exit node (A) returns `["close_a", "open_b"]`, we link exit node (A) to the entry node of method `close_a`, and exit node (A) to the entry node of method `open_b`.

## 3.2 Method behavior extraction

In this section we formalize how our tool infers the behavior of each method, described as a regular expression. *The formalization included in this paper, along with the theoretical results presented, are fully mechanized using the Coq proof assistant.* The syntax of the source language is an abstraction of MicroPython, that captures the control flow of the program and function calls — our input language ignores the intermediate values being calculated. The syntax of the source language, its semantics, and the behavior inference is given in Figure 4.
**Syntax.** The syntax of a program $p$ consists of: $f()$ to encode a method call (discarding the arguments); `skip` represents any MicroPython instruction that is of no interest to the analysis; `return` to return a value, where the actual value being returned is ignored at this stage of the analysis; $p_1; p_2$ sequences program $p_1$ followed by program $p_2$; `if`$(\star)$ $\{p_1\}$ `else` $\{p_2\}$ represents a non-deterministic choice, as the condition is not represented; `loop`$(\star)$ $\{p\}$ represents a loop that runs $p$ an unknown number of iterations. *Supported Python constructs:* our analysis represents `for` and `while` as a loop, `match` and `if` as a conditional; our analysis does not model Python exceptions.
**Semantics.** We give semantics to our syntax in terms of the sequences of labels that a program outputs. The judgment $s \vdash l \in p$ denotes that trace $l$ (a sequence

**Syntax**

$$p \ ::= \ f() \ \mid \ \texttt{skip} \ \mid \ \texttt{return} \ \mid \ p;p \ \mid \ \texttt{if}(\star)\ \{p\}\ \texttt{else}\ \{p\} \ \mid \ \texttt{loop}(\star)\ \{p\}$$

$$s \ ::= \ \texttt{0} \ \mid \ \texttt{R}$$

**Semantics** $\boxed{s \vdash l \in p}$

$$\text{CALL}\ \frac{}{\texttt{0} \vdash [f] \in f()} \qquad \text{SKIP}\ \frac{}{\texttt{0} \vdash [] \in \texttt{skip}} \qquad \text{RETURN}\ \frac{}{\texttt{R} \vdash [] \in \texttt{return}} \qquad \text{SEQ-1}\ \frac{\texttt{R} \vdash l \in p_1}{\texttt{R} \vdash l \in p_1; p_2}$$

$$\text{SEQ-2}\ \frac{\texttt{0} \vdash l_1 \in p_1 \qquad s \vdash l_2 \in p_2}{s \vdash l_1 \cdot l_2 \in p_1; p_2} \qquad \text{IF-1}\ \frac{s \vdash l \in p_1}{s \vdash l \in \texttt{if}(\star)\ \{p_1\}\ \texttt{else}\ \{p_2\}}$$

$$\text{IF-2}\ \frac{s \vdash l \in p_2}{s \vdash l \in \texttt{if}(\star)\ \{p_1\}\ \texttt{else}\ \{p_2\}} \qquad \text{LOOP-1}\ \frac{}{\texttt{0} \vdash [] \in \texttt{loop}(\star)\ \{p\}} \qquad \text{LOOP-2}\ \frac{\texttt{R} \vdash l \in p}{\texttt{R} \vdash l \in \texttt{loop}(\star)\ \{p\}}$$

$$\text{LOOP-3}\ \frac{\texttt{0} \vdash l_1 \in p \qquad s \vdash l_2 \in \texttt{loop}(\star)\ \{p\}}{s \vdash l_1 \cdot l_2 \in \texttt{loop}(\star)\ \{p\}}$$

**Behavior inference** $\boxed{[\![p]\!] = (r, s)}$ $\boxed{\text{infer}(p) = r}$

$$[\![f()]\!] = (f, \emptyset) \qquad [\![\texttt{skip}]\!] = (\epsilon, \emptyset) \qquad [\![\texttt{return}]\!] = (\emptyset, \{\epsilon\})$$

$$[\![p_1; p_2]\!] = (r_1 \cdot r_2, \{r_1 \cdot r \mid r \in s_2\} \cup s_1)$$

$$[\![\texttt{if}(\star)\ \{p_1\}\ \texttt{else}\ \{p_2\}]\!] = (r_1 + r_2, s_1 \cup s_2)$$

$$[\![\texttt{loop}(\star)\ \{p_1\}]\!] = (r_1^\star, \{r_1^\star \cdot r \mid r \in s_1\})$$

$$\text{where } [\![p_1]\!] = (r_1, s_1) \text{ and } [\![p_2]\!] = (r_2, s_2)$$

$$\text{infer}(p) = r_1 + r_1' + \cdots + r_n' \text{ where } [\![p]\!] = (r_1, s_1) \wedge s_1 = \{r_1', \ldots, r_n'\}$$

Figure 4: Extracting the behavior from imperative code

of labels $f$) is output by program $p$ when the program has a certain status $s$ (which is either $\texttt{0}$ for ongoing, or $\texttt{R}$ for returned). Our intuition behind using the term "returned" is in the sense that no other terms of the trace can be in a returned trace, whereas a "ongoing" trace can be sequenced further. We denote a sequence comma separated between brackets, *e.g.,* a sequence with digits 0, 1, and 2 is denoted by $[0, 1, 2]$. The usual sequence concatenation is denoted by $l_1 \cdot l_2$. Rule CALL states that trace $[f]$ is in program $f()$ when the program is still ongoing ($\texttt{0}$). Rule SKIP states that empty trace [] is in program $\texttt{skip}$ when the program is still ongoing ($\texttt{0}$). Rule RETURN states that empty trace [] is in program $\texttt{skip}$ when the program has returned ($\texttt{R}$). The rules are that of sequence. Rule SEQ-1 states that if a certain trace $l$ is in $p_1$ and that trace has returned (due to a return), then program $p_1; p_2$ also outputs trace $l$. Rule SEQ-2 states what happens if trace $l_1$ of $p_1$ is ongoing ($\texttt{0}$), then we can prepend $l_1$ to any trace $l_2$ from $p_2$ in $p_1; p_2$. Rule IF-1 and IF-2 state that if either branch $p_1$ or branch $p_2$ output a trace $l$ under status $s$, then program $\texttt{if}(\star)\ \{p_1\}\ \texttt{else}\ \{p_2\}$ outputs a trace $l$ under status $s$. Finally, the loop is governed by three rules. Rule LOOP-1 (akin to Rule SKIP) states that a loop can terminate and in that case we have an empty trace [] and the status is ongoing. Rule LOOP-2 (akin to Rule SEQ-1) captures the case where the loop body $p$ outputs trace $l$ and issues a return. Rule LOOP-3 (akin to Rule SEQ-2) captures the case where the loop body $p$ outputs a trace $l_1$ and the computation continues by ongoing $\texttt{loop}(\star)\ \{p\}$ with a trace $l_2$.

*Example 1.* The trace $[a, c, a, c]$ is in the following program, which exercises a trace that yields from a loop that runs 2 iterations without an early return.

$$\texttt{0} \vdash [a, c, a, c] \in \texttt{loop}(\star)\ \{a(); \texttt{if}(\star)\ \{b(); \texttt{return}\}\ \texttt{else}\ \{c()\}\}$$

*Example 2.* The trace $[a, c, a, b]$ is in the same program, which highlights the case where the loop runs for one iteration and in the second iteration returns.

$$\texttt{R} \vdash [a, c, a, b] \in \texttt{loop}(\star)\ \{a(); \texttt{if}(\star)\ \{b(); \texttt{return}\}\ \texttt{else}\ \{c()\}\}$$

**Definition 1** (Behavior)**.** *Let* $L(p) = \{l \mid s \vdash l \in p\}$ *denote the* behavior *of program* $p$*.*

**Behavior inference.** We now introduce the process of extracting the program's behavior as a regular expression. Let $r$ denote a regular expression, defined as follows.

$$r \ ::= \ \epsilon \mid \emptyset \mid f \mid r \cdot r \mid r + r \mid r^\star$$

where $\epsilon$ denotes the empty string, $\emptyset$ denotes the empty set, $f$ denotes a set only containing $f$, $r \cdot r$ denotes set concatenation, $r + r$ denotes set union, and $r^\star$ denotes the Kleene-star operator. Function $[\![p]\!] = (r, s)$ takes a program $p$ and outputs a pair that holds a regular expression $r$ of the ongoing behavior and a set of returned behaviors $s$, where $s$ is a finite set of regular expressions. Intuitively, the inference's output captures the two kinds of status: the first component $r$ represents $\texttt{0} \vdash l \in p$, and the second component $s$ represents $\texttt{R} \vdash l \in p$. The case

for $f()$ yields regular expression $f$ and no returned behaviors. The case for `skip` yields $\epsilon$ and no returned behaviors. The case for `return` is more interesting: the function yields $\emptyset$ so that no behavior can ensue the return, and the empty string in the set of returned behaviors arises from $\mathtt{R} \vdash [] \in \mathtt{return}$ in Rule RET. The case for $p_1; p_2$ the $s_1$ on the right-hand side of the output is due to any early return of $p_1$ (Rule SEQ-1); the $r_1 \cdot r_2$ on the left-hand side and $\{r_1 \cdot r \mid r \in s_2\}$ capture the behavior of program $p_1$ without early returns (Rule SEQ-2). The conditional is represents the union of behaviors. The case for loop returns the Kleene-star of the left-hand side the behavior $r_1$ of $p$ without early returns; on the right-hand side the function prepends zero or more iterations of $r_1$ followed by the behavior of each early return $r$. Finally, infer($p$) merges the running and all the halted behaviors.

*Example 3.* We can now translate the program used in Examples 1 and 2.

$$[\![\mathtt{loop}(\star)\ \{a(); \mathtt{if}(\star)\ \{b(); \mathtt{return}\}\ \mathtt{else}\ \{c()\}\}]\!] =$$

$$\left(\left(a \cdot ((b \cdot \emptyset) + c)\right)^\star, \left\{\left(a \cdot ((b \cdot \emptyset) + c)\right)^\star \cdot a \cdot b\right\}\right)$$

**Theorem 1** (Soundness)**.** *If $l \in L(p)$, then $l \in \text{infer}(p)$.*

*Proof.* To establish this result, we must first show the following lemma (1). If $[\![p]\!] = (r, s)$ and $\mathtt{0} \vdash l \in p$, then $l \in r$. The proof follows by induction on the derivation of $\mathtt{0} \vdash l \in p$.

Next, we show (2) that if $[\![p]\!] = (r, s)$ and $\mathtt{R} \vdash l \in p$, then there exists an $r' \in s$ and $l \in r'$. The proof follows by induction on the derivation of $\mathtt{R} \vdash l \in p$. The interesting cases are Rule SEQ-1 and Rule LOOP-3, which require lemma (1).

Our goal is to show that $l \in \text{infer}(p) = l \in r + r'_1 + \cdots + r'_n$ where $s = \{r'_1, \ldots, r'_n\}$ and $[\![p]\!] = (r, s)$. Given our assumption $l \in L(p)$, then there are two cases to consider. Case $\mathtt{0} \vdash l \in p$, then we apply (1), and obtain that $l \in r$, hence $l \in r + r'_1 + \cdots + r'_n$.

Case $\mathtt{R} \vdash l \in p$, then we apply (2) to know that there exists an $r' \in s$ and $l \in r'$. Given assumption $l \in r'$, then by induction on the structure of $s$ we can derive that $l \in r + r'_1 + \cdots + r'_n$. □

**Theorem 2** (Completeness)**.** *If $l \in \text{infer}(p)$, then $l \in L(p)$.*

*Proof.* The proof follows the same structure of Theorem 1. We show (1): If $[\![p]\!] = (r, s)$ and $l \in r$, then $\mathtt{0} \vdash l \in p$. The proof follows by induction on the derivation of $l \in r$.

Next, we show (2) that if $[\![p]\!] = (r, s)$, $r' \in s$, and $l \in r'$, then $\mathtt{R} \vdash l \in p$. The proof follows by induction on the structure of $p$. The interesting cases are sequence, and loop, both of which require the use of (1). To conclude the case for loop we also require an auxiliary result: if $\mathtt{0} \vdash l_1 \in \mathtt{loop}(\star)\ \{p\}$ and $s \vdash l_2 \in p$, then $s \vdash l_1 \cdot l_2 \in \mathtt{loop}(\star)\ \{p\}$.

We have that $l \in \text{infer}(p) \equiv l \in r + r'_1 + \cdots + r'_n$ where $s = \{r'_1, \ldots, r'_n\}$ and $[\![p]\!] = (r, s)$. We must show that $\exists s$ such that $s \vdash l \in p$. By induction on the output of infer($p$) we can show that either $l \in r$ or there exist $r'$ such that $r' \in s$ and $l \in r'$. For the former we use (1) and for the latter we use (2). □

Our main theoretical result is the correctness of our extraction process (function infer($p$), which states that the behavior of a program is a regular language.

**Corollary 1.** *We have that $L(p)$ is a regular language.*

*Proof.* $L(p)$ is regular since infer($p$) recognizes $L(p)$. □

# 4 Related work

There have been different approaches to model check programming languages. Regarding C++, several tools focus on memory safety issues and specific features such as exception handling [41, 8, 14, 6]. Some tools analyze an intermediate language, *e.g.,* LLVM [23, 52, 1] but then face the challenge of losing context information. In [5], authors present a comparative evaluation of fully automatic software verifiers for C and Java [35, 15, 30, 44]. In particular, Java PathFinder [27] is a very well-known tool in this domain. For Lustre: JKind [24, 30]. For Python: MSVL [49]. Domain-specific languages that can be verified for correctness include P [20, 21], Rebeca [50], and synchronous reactive languages[4, 29, 3]

Utilizing Finite State Machines (FSMs) for program behavior modeling is a prevalent approach. For instance, in [54], the authors suggest a technique for extracting finite-state models of object-oriented class interfaces automatically and they check whether the program exhibits the expected behavior. VeriSolid [43, 40] applies formal methods to verify smart contracts specified as transition-systems, and includes a visualization tool. JavaBIP [9] is a framework that uses annotations directly on Java code in order to coordinate existing concurrent software components.

Typestates [51, 25] refine the concept of type with information about which operations can be used in a particular context. Multiple authors apply typestates to general-purpose programming languages [36, 22, 53, 11]. Shelley explores a similar notion but from a modeling checking perspective; moreover typestates are based on state-change, rather than on call ordering constraints.

# 5 Conclusion and Future Work

In this paper, we present the inference of Shelley specifications from MicroPython classes, which are used in the context of model checking. The inference process consists of three steps: method dependency extraction, method behavior extraction, and method invocation analysis. Our main contribution is formalizing and establishing the correctness of method behavior extraction that specifies the behavior of a small imperative calculus as a regular language.

**Future work.** Shelley delegates the actual model checking to NuSMV [13], by implementing a translation from a nondeterministic finite automaton (NFA) into a NuSMV model. Our approach is essentially to encode a regular-language as a $\omega$-regular language. We would like to evaluate other approaches that work directly in regular-languages, such as [19] and [33].

**Acknowledgements** This material is based upon work supported by the National Science Foundation under Grant No. 2204986. This work was supported by FCT through scholarships SFRH/BD/131418/2017 and SFRH/BI/153747/2019, and the LASIGE Research Unit, ref. UIDB/00408/2020 and ref. UIDP/00408/2020.

# References

[1] Zuzana Baranová, Jiri Barnat, Katarína Kejstová, Tadeás Kucera, Henrich Lauko, Jan Mrázek, Petr Rockai, and Vladimír Still. Model checking of C and C++ with DIVINE 4. In *ATVA*, volume 10482 of *LNCS*, pages 201–207. Springer, 2017.

[2] Andreas Bauer, Martin Leucker, and Christian Schallhart. Comparing LTL semantics for runtime verification. *Journal of Logic and Computation*, 20(3):651–674, 2010.

[3] Albert Benveniste, Paul Le Guernic, and Christian Jacquemot. Synchronous Programming with Events and Relations: the SIGNAL Language and Its Semantics. *Sci. Comput. Program.*, 16(2):103–149, 1991.

[4] Gérard Berry and Laurent Cosserat. The ESTEREL Synchronous Programming Language and its Mathematical Semantics. In *SC*, volume 197 of *LNCS*, pages 389–448. Springer, 1984.

[5] Dirk Beyer. Automatic verification of C and java programs: SV-COMP 2019. In *TACAS*, volume 11429 of *LNCS*, pages 133–155. Springer, 2019.

[6] Dirk Beyer, Thomas A. Henzinger, Ranjit Jhala, and Rupak Majumdar. The software model checker blast. *Int. J. Softw. Tools Technol. Transf.*, 9(5-6):505–525, 2007.

[7] Dirk Beyer and M. Erkan Keremoglu. CPACHECKER: A tool for configurable software verification. In *CAV*, page 184–190, Berlin, Heidelberg, 2011. Springer-Verlag.

[8] Nicolas Blanc, Alex Groce, and Daniel Kroening. Verifying C++ with STL containers via predicate abstraction. In *ASE*, pages 521–524. ACM, 2007.

[9] Simon Bliudze, Anastasia Mavridou, Radoslaw Szymanek, and Alina Zolotukhina. Exogenous coordination of concurrent software components with JavaBIP. *Softw. Pract. Exp.*, 47(11):1801–1836, 2017.

[10] Ivan Bocic and Tevfik Bultan. Symbolic model extraction for web application verification. In *ICSE*, pages 724–734. IEEE / ACM, 2017.

[11] Mario Bravetti, Adrian Francalanza, Iaroslav Golovanov, Hans Hüttel, Mathias Jakobsen, Mikkel Kettunen, and António Ravara. Behavioural types for memory and method safety in a core object-oriented language. In *APLAS*, volume 12470 of *LNCS*, pages 105–124. Springer, 2020.

[12] Olav Bunte, Jan Friso Groote, Jeroen JA Keiren, Maurice Laveaux, Thomas Neele, Erik P de Vink, Wieger Wesselink, Anton Wijs, and Tim AC Willemse. The mCRL2 toolset for analysing concurrent systems: improvements in expressivity and usability. In *TACAS*, pages 21–39. Springer, 2019.

[13] Alessandro Cimatti, Edmund M. Clarke, Enrico Giunchiglia, Fausto Giunchiglia, Marco Pistore, Marco Roveri, Roberto Sebastiani, and Armando Tacchella. NuSMV 2: An opensource tool for symbolic model checking. In *CAV*, volume 2404 of *LNCS*, pages 359–364. Springer, 2002.

[14] Edmund M. Clarke, Daniel Kroening, and Flavio Lerda. A tool for checking ANSI-C programs. In *TACAS*, volume 2988 of *LNCS*, pages 168–176. Springer, 2004.

[15] Lucas C. Cordeiro, Pascal Kesseli, Daniel Kroening, Peter Schrammel, and Marek Trtík. JBMC: A bounded model checking tool for verifying java bytecode. In *CAV*, volume 10981 of *LNCS*, pages 183–190. Springer, 2018.

[16] Carlos Mão de Ferro, Tiago Cogumbreiro, and Francisco Martins. Formalizing Model Inference of MicroPython, may 2023.

[17] Carlos Mão de Ferro, Tiago Cogumbreiro, and Francisco Martins. Shelley, a framework for model checking call ordering on hierarchical systems. In *COORDINATION*, LNCS. Springer, 2023.

[18] Carlos Mão de Ferro, Tiago Cogumbreiro, and Francisco Martins. Shelley: a framework for model checking call ordering on hierarchical systems, may 2023.

[19] Giuseppe De Giacomo and Moshe Y. Vardi. Linear temporal logic and linear dynamic logic on finite traces. In *IJCAI*, page 854–860. AAAI Press, 2013.

[20] Ankush Desai et al. P: Safe Asynchronous Event-driven Programming. In *PLDI*, pages 321–332. ACM, 2013.

[21] Ankush Desai, Amar Phanishayee, Shaz Qadeer, and Sanjit A. Seshia. Compositional programming and testing of dynamic distributed systems. *Proc. ACM Program. Lang.*, 2(OOPSLA), 2018.

[22] José Duarte and António Ravara. Retrofitting Typestates into Rust. In *SBLP*, pages 83–91. ACM, 2021.

[23] Stephan Falke, Florian Merz, and Carsten Sinz. The bounded model checker LLBMC. In *ASE*, pages 706–709. IEEE, 2013.

[24] Andrew Gacek, John Backes, Mike Whalen, Lucas Wagner, and Elaheh Ghassabani. The JKind model checker. In *CAV*, pages 20–27. Springer, 2018.

[25] Ronald Garcia, Éric Tanter, Roger Wolff, and Jonathan Aldrich. Foundations of typestate-oriented programming. *ACM Trans. Program. Lang. Syst.*, 36(4), 2014.

[26] Damien George. MicroPython, 2022.

[27] Dimitra Giannakopoulou and Corina S. Pasareanu. Interface generation and compositional verification in JavaPathfinder. In *FASE*, volume 5503 of *LNCS*, pages 94–108. Springer, 2009.

[28] Fernando Silvano Goncalves, David Pereira, Eduardo Tovar, and Leandro Buss Becker. Formal verification of AADL models using UPPAAL. In *SBESC*, pages 117–124. IEEE Computer Society, 2017.

[29] N. Halbwachs, P. Caspi, P. Raymond, and D. Pilaud. The synchronous data flow programming language LUSTRE. *Proceedings of the IEEE*, 79(9):1305–1320, 1991.

[30] John Hatcliff and Matthew B. Dwyer. Using the Bandera tool set to model-check properties of concurrent Java software. In *CONCUR*, volume 2154 of *LNCS*, pages 39–58. Springer, 2001.

[31] Gerard J. Holzmann. The model checker SPIN. *IEEE Trans. Softw. Eng.*, 23(5):279–295, 1997.

[32] Gerard J. Holzmann and Margaret H. Smith. Software model checking: extracting verification models from source code. *Softw. Test. Verification Reliab.*, 11(2):65–79, 2001.

[33] Samuel Huang and Rance Cleaveland. A tableau construction for finite linear-time temporal logic. *Journal of Logical and Algebraic Methods in Programming*, 125:100743, 2022.

[34] Joel Huselius, Johan Kraft, Hans Hansson, and Sasikumar Punnekkat. Evaluating the Quality of Models Extracted from Embedded Real-Time Software. In *ECBS*, pages 577–585. IEEE Computer Society, 2007.

[35] Temesghen Kahsai, Philipp Rümmer, Huascar Sanchez, and Martin Schäf. Jayhorn: A framework for verifying java programs. In *CAV*, volume 9779 of *LNCS*, pages 352–358. Springer, 2016.

[36] Dimitrios Kouzapas, Ornela Dardha, Roly Perera, and Simon J. Gay. Typechecking protocols with Mungo and StMungo: A session type toolchain for Java. *Sci. Comput. Program.*, 155:52–75, 2018.

[37] Kim Guldstrand Larsen, Paul Pettersson, and Wang Yi. UPPAAL in a nutshell. *Int. J. Softw. Tools Technol. Transf.*, 1(1-2):134–152, 1997.

[38] Anastasia Mavridou, Hamza Bourbouh, Pierre-Loïc Garoche, Dimitra Giannakopoulou, Thomas Pressburger, and Johann Schumann. Bridging the Gap Between Requirements and Simulink Model Analysis. In *REFSQ*, volume 2584 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020.

[39] Anastasia Mavridou, Hamza Bourbouh, Dimitra Giannakopoulou, Thomas Pressburger, Mohammad Hejase, Pierre-Loïc Garoche, and Johann Schumann. The Ten Lockheed Martin Cyber-Physical Challenges: Formalized, Analyzed, and Explained. In *RE*, pages 300–310. IEEE, 2020.

[40] Anastasia Mavridou, Aron Laszka, Emmanouela Stachtiari, and Abhishek Dubey. VeriSolid: Correct-by-Design Smart Contracts for Ethereum. *CoRR*, abs/1901.01292, 2019.

[41] Felipe R. Monteiro, Mikhail R. Gadelha, and Lucas C. Cordeiro. Model checking C++ programs. *Softw. Test. Verification Reliab.*, 32(1), 2022.

[42] Shiva Nejati, Khouloud Gaaloul, Claudio Menghi, Lionel C. Briand, Stephen Foster, and David Wolfe. Evaluating model testing and model checking for finding requirements violations in simulink models. In *FSE*, pages 1015–1025. ACM, 2019.

[43] Keerthi Nelaturu, Anastasia Mavridou, Andreas G. Veneris, and Aron Laszka. Verified Development and Deployment of Multiple Interacting Smart Contracts with VeriSolid. In *ICBC*, pages 1–9. IEEE, 2020.

[44] Yannic Noller, Corina S. Pasareanu, Aymeric Fromherz, Xuan-Bach Dinh Le, and Willem Visser. Symbolic pathfinder for SV-COMP - (competition contribution). In *TACAS*, volume 11429 of *LNCS*, pages 239–243. Springer, 2019.

[45] Jevitha K. P., Swaminathan Jayaraman, Bharat Jayaraman, and M. Sethumadhavan. Finite-state model extraction and visualization from java program execution. *Softw. Pract. Exp.*, 51(2):409–437, 2021.

[46] Pavel Parízek, Frantisek Plasil, and Jan Kofron. Model Checking of Software Components: Combining Java PathFinder and Behavior Protocol Model Checker. In *SEW*, pages 133–141. IEEE, 2006.

[47] André Santos, Alcino Cunha, and Nuno Macedo. Static-Time Extraction and Analysis of the ROS Computation Graph. In *IRC*, pages 62–69. IEEE, 2019.

[48] Maike Schwammberger and Verena Klös. From Specification Models to Explanation Models: An Extraction and Refinement Process for Timed Automata. In *FMAS*, volume 371 of *EPTCS*, pages 20–37, 2022.

[49] Xinfeng Shu, Fengyun Gao, Weiran Gao, Lili Zhang, Xiaobing Wang, and Liang Zhao. Model Checking Python Programs with MSVL. In *SOFL*, volume 12028 of *LNCS*, pages 205–224. Springer, 2019.

[50] Marjan Sirjani and Mohammad Mahdi Jaghoori. Ten Years of Analyzing Actors: Rebeca Experience. In *Essays Dedicated to Carolyn Talcott on the Occasion of Her 70th Birthday*, volume 7000 of *LNCS*, pages 20–56. Springer, 2011.

[51] R E Strom and S Yemini. Typestate: A Programming Language Concept for Enhancing Software Reliability. *IEEE Trans. Softw. Eng.*, 12(1):157–171, 1986.

[52] S. Thompson and G. Brat. Verification of c++ flight software with the mcp model checker. In *IEEE Aerospace Conference*, pages 1–9, 2008.

[53] A. Laura Voinea, Ornela Dardha, and Simon J. Gay. Typechecking Java Protocols with [St]Mungo. In *FORTE*, volume 12136 of *LNCS*, pages 208–224. Springer, 2020.

[54] John Whaley, Michael C. Martin, and Monica S. Lam. Automatic extraction of object-oriented component interfaces. In *ISSTA*, pages 218–228. ACM, 2002.