ACSL: ANSI/ISO C Specification Language
Version 1.9

Software Analyzers

ACSL: ANSI/ISO C Specification Language
Version 1.9
ACSL: ANSI/ISO C Specification Language

Version 1.9

Patrick Baudin¹, Pascal Cuoq¹, Jean-Christophe Filliâtre⁴,³, Claude Marché⁴,³, Benjamin Monate¹, Yannick Moy²,³,⁴, Virgile Prevosto¹

¹ CEA LIST, Software Reliability Laboratory, Saclay, F-91191
² France Télécom, Lannion, F-22307
³ INRIA Saclay - Île-de-France, ProVal, Orsay, F-91893
⁴ LRI, Univ Paris-Sud, CNRS, Orsay, F-91405

©2009-2013 CEA LIST and INRIA

This work has been supported by the ANR project CAT (ANR-05-RNTL-0030x), by the ANR CIFRE contract 2005/973, by the ANR project U3CAT (08-SEGI-021-xx), and ANR PICF project DEVICE-Soft (2009-CARN-006-01).
# Contents

1 Introduction

1.1 Organization of this document ........................................... 11
1.2 Generalities about Annotations .......................................... 11
  1.2.1 Kinds of annotations .................................................. 12
  1.2.2 Parsing annotations in practice .................................... 12
  1.2.3 About preprocessing ................................................. 13
  1.2.4 About keywords ...................................................... 13
1.3 Notations for grammars ................................................... 13

2 Specification language

2.1 Lexical rules .................................................................. 15
2.2 Logic expressions .......................................................... 16
  2.2.1 Operators precedence ................................................. 19
  2.2.2 Semantics ............................................................... 19
  2.2.3 Typing ................................................................ 20
  2.2.4 Integer arithmetic and machine integers ......................... 20
  2.2.5 Real numbers and floating point numbers ....................... 23
  2.2.6 C arrays and pointers ............................................... 25
  2.2.7 Structures, Unions and Arrays in logic ......................... 26
  2.2.8 String literals ......................................................... 28
2.3 Function contracts .......................................................... 28
  2.3.1 Built-in constructs `\old` and `\result` ............................. 29
  2.3.2 Simple function contracts ............................................ 30
  2.3.3 Contracts with named behaviors ................................... 31
  2.3.4 Memory locations and sets of terms ............................... 34
  2.3.5 Default contracts, multiple contracts ............................. 35
2.4 Statement annotations ..................................................... 36
  2.4.1 Assertions ............................................................ 36
  2.4.2 Loop annotations ..................................................... 36
2.4.3 Built-in construct \at ........................................... 41
2.4.4 Statement contracts ........................................... 43
2.5 Termination ....................................................... 44
  2.5.1 Integer measures ........................................... 45
  2.5.2 General measures ........................................... 45
  2.5.3 Recursive function calls ................................... 46
  2.5.4 Non-terminating functions ................................. 47
2.6 Logic specifications ............................................ 47
  2.6.1 Predicate and function definitions ......................... 48
  2.6.2 Lemmas ...................................................... 49
  2.6.3 Inductive predicates ....................................... 49
  2.6.4 Axiomatic definitions ...................................... 50
  2.6.5 Polymorphic logic types .................................. 51
  2.6.6 Recursive logic definitions ................................ 51
  2.6.7 Higher-order logic constructions ......................... 52
  2.6.8 Concrete logic types ...................................... 53
  2.6.9 Hybrid functions and predicates .......................... 54
  2.6.10 Memory footprint specification: \texttt{reads} clause ....... 55
  2.6.11 Specification Modules .................................... 56
2.7 Pointers and physical addressing ............................... 57
  2.7.1 Memory blocks and pointer dereferencing ................ 57
  2.7.2 Separation .................................................. 59
  2.7.3 Dynamic allocation and deallocation ....................... 59
2.8 Sets as first-class values ...................................... 63
2.9 Abrupt termination ............................................. 64
2.10 Dependencies information ..................................... 66
2.11 Data invariants ............................................... 67
  2.11.1 Semantics .................................................. 68
  2.11.2 Model variables and model fields ......................... 70
2.12 Ghost variables and statements ............................... 72
  2.12.1 Volatile variables ........................................ 75
2.13 Undefined values, dangling pointers ......................... 76
  2.13.1 Initialization ............................................... 76
  2.13.2 Dangling pointers ......................................... 77
This is the version 1.9 of ACSL design. Several features may still evolve in the future. In particular, some features in this document are considered experimental, meaning that their syntax and semantics is not already fixed. These features are marked with Experimental. They must also be considered as advanced features, which are not supposed to be useful for a basic use of that specification language.

Acknowledgements

We gratefully thank all the people who contributed to this document: Sylvie Boldo, Jean-Louis Colaço, Pierre Crégut, David Delmas, Catherine Dubois, Stéphane Duprat, Arnaud Gotlieb, Philippe Herrmann, Thierry Hubert, Dillon Pariente, Pierre Rousseau, Julien Signoles, Jean Souyris, Asma Tafat.
Chapter 1

Introduction

This document is a reference manual for ACSL. ACSL is an acronym for “ANSI/ISO C Specification Language”. This is a Behavioral Interface Specification Language (BISL) implemented in the FRAMA-C framework. It aims at specifying behavioral properties of C source code. The main inspiration for this language comes from the specification language of the CADUCEUS tool [9, 10] for deductive verification of behavioral properties of C programs. The specification language of Caduceus is itself inspired from the Java Modeling Language (JML [18]) which aims at similar goals for Java source code: indeed it aims both at runtime assertion checking and static verification using the ESC/Java2 tool [14], where we aim at static verification and deductive verification (see Appendix A.2 for a detailed comparison between ACSL and JML).

Going back further in history, JML design was guided by the general design-by-contract principle proposed by Bertrand Meyer, who took his own inspiration from the concepts of preconditions and postconditions on a routine, going back at least to Dijkstra, Floyd and Hoare in the late 60’s and early 70’s, and originally implemented in the EIFFEL language.

In this document, we assume that the reader has a good knowledge of the ISO C programming language [13, 12].

1.1 Organization of this document

In this preliminary chapter we introduce some definitions and vocabulary, and discuss generalities about this specification language. Chapter 2 presents the specification language itself. Chapter 3 presents additional information about libraries of specifications. Appendix A provides specific hindsight over type-checking ACSL annotations, the relation between ACSL and JML, and specification templates. A detailed table of contents is given on page 5. A glossary is given in Appendix A.1.

1.2 Generalities about Annotations

In this document, we consider that specifications are given as annotations in comments written directly in C source files, so that source files remain compilable. Those comments must start by /*@ or //@ and end as usual in C.

In some contexts, it is not possible to modify the source code. It is strongly recommended that a tool which implements ACSL specifications provides technical means to store annotations
separately from the source. It is not the purpose of this document to describe such means. Nevertheless, some of the specifications, namely those at a global level, can be given in separate files: logical specifications can be imported (see Section 2.6.11) and a function contract can be attached to a copy of the function profile (see Section 2.3.5).

1.2.1 Kinds of annotations

- Global annotations:
  - function contract: such an annotation is inserted just before the declaration or the definition of a function. See section 2.3.
  - global invariant: this is allowed at the level of global declarations. See section 2.11.
  - type invariant: this allows to declare both structure or union invariants, and invariants on type names introduced by typedef. See section 2.11.
  - logic specifications: definitions of logic functions or predicates, lemmas, axiomatizations by declaration of new logic types, logic functions, predicates with axioms they satisfy. Such an annotation is placed at the level of global declarations. See section 2.6

- Statement annotations:
  - assertion: these are allowed everywhere a C label is allowed, or just before a block closing brace. See section 2.4.1.
  - loop annotation (invariant, variant, assign clauses): is allowed immediately before a loop statement: for, while, do...while. See Section 2.4.2.
  - statement contract: very similar to a function contract, and placed before a statement or a block. Semantical conditions must be checked (no goto going inside, no goto going outside). See Section 2.4.4.
  - ghost code: regular C code, only visible from the specifications, that is only allowed to modify ghost variables. See section 2.12. This includes ghost braces for enclosing blocks.

1.2.2 Parsing annotations in practice

In JML, parsing is done by simply ignoring //@,/*@ and */ at the lexical analysis level. This technique could modify the semantics of the code, for example:

```c
1     return x /@ +1 */ ;
```

In our language, this is forbidden. Technically, the current implementation of Frama-C isolates the comments in a first step of syntax analysis, and then parses a second time. Nevertheless, the grammar and the corresponding parser must be carefully designed to avoid interaction of annotations with the code. For example, in code such as

```c
1     if (c) //@ assert P;
2         c=1;
```

the statement c=1 must be understood as the branch of the if. This is ensured by the grammar below, saying that assert annotations are not statements themselves, but attached to the statement that follows, like C labels.
1.2.3 About preprocessing

This document considers C source after preprocessing. Tools must decide how they handle preprocessing (what to do with annotations, whether macro substitution should be performed, etc.)

1.2.4 About keywords

Additional keywords of the specification language start with a backslash, if they are used in position of a term or a predicate (which are defined in the following). Otherwise they do not start with a backslash (like \texttt{ensures}) and they remain valid identifiers.

1.3 Notations for grammars

In this document, grammar rules are given in BNF form. In grammar rules, we use extra notations $e^*$ to denote repetition of zero, one or more occurrences of $e$, $e^+$ for repetition of one or more occurrences of $e$, and $e?$ for zero or one occurrence of $e$. For the sake of simplicity, we only describe annotations in the usual \texttt{/*@ ... */} style of comments. One-line annotations in \texttt{/**< ... */} comments are alike.
Chapter 2

Specification language

2.1 Lexical rules

Lexical structure mostly follows that of ANSI/ISO C. A few differences must be noted.

- The at sign (@) is a blank character, thus equivalent to a space character.
- Identifiers may start with the backslash character (\).
- Some UTF8 characters may be used in place of some constructs, as shown in the following table:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>ASCII Code</th>
<th>Hex Code</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&gt;=</code></td>
<td><code>&gt;</code></td>
<td>0x2265</td>
</tr>
<tr>
<td><code>&lt;=</code></td>
<td><code>&lt;</code></td>
<td>0x2264</td>
</tr>
<tr>
<td><code>&gt;</code></td>
<td><code>&gt;</code></td>
<td>0x003E</td>
</tr>
<tr>
<td><code>&lt;</code></td>
<td><code>&lt;</code></td>
<td>0x003C</td>
</tr>
<tr>
<td><code>!=</code></td>
<td><code>≠</code></td>
<td>0x2262</td>
</tr>
<tr>
<td><code>==</code></td>
<td><code>≡</code></td>
<td>0x2261</td>
</tr>
<tr>
<td><code>=&gt;</code></td>
<td><code>⇒</code></td>
<td>0x21D2</td>
</tr>
<tr>
<td><code>&lt;=&gt;</code></td>
<td><code>⇔</code></td>
<td>0x21D4</td>
</tr>
<tr>
<td><code>&amp;&amp;</code></td>
<td><code>∧</code></td>
<td>0x2227</td>
</tr>
<tr>
<td>`</td>
<td></td>
<td>`</td>
</tr>
<tr>
<td><code>^^</code></td>
<td><code>⊕</code></td>
<td>0x22BB</td>
</tr>
<tr>
<td><code>!</code></td>
<td><code>¬</code></td>
<td>0x00AC</td>
</tr>
<tr>
<td><code>-</code></td>
<td><code>-</code></td>
<td>0x2212</td>
</tr>
</tbody>
</table>

  - Comments can be put inside ACSL annotations. They use the C++ format, i.e. begin with `//` and extend to the end of current line.
### CHAPTER 2. SPECIFICATION LANGUAGE

#### Figure 2.1: Grammar of terms

<table>
<thead>
<tr>
<th>Grammar</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>literal ::=</code></td>
<td>boolean constants</td>
</tr>
<tr>
<td>`</td>
<td>integer</td>
</tr>
<tr>
<td>`</td>
<td>real</td>
</tr>
<tr>
<td>`</td>
<td>string</td>
</tr>
<tr>
<td>`</td>
<td>character</td>
</tr>
<tr>
<td><code>bin-op ::=</code></td>
<td>boolean operations</td>
</tr>
<tr>
<td>`</td>
<td>+</td>
</tr>
<tr>
<td>`</td>
<td>==</td>
</tr>
<tr>
<td>`</td>
<td>&amp; &amp;</td>
</tr>
<tr>
<td>`</td>
<td>&amp;</td>
</tr>
<tr>
<td><code>unary-op ::=</code></td>
<td>unary plus and minus</td>
</tr>
<tr>
<td>`</td>
<td>+</td>
</tr>
<tr>
<td>`</td>
<td>!`</td>
</tr>
<tr>
<td>`</td>
<td>~`</td>
</tr>
<tr>
<td>`</td>
<td>*`</td>
</tr>
<tr>
<td>`</td>
<td>&amp;`</td>
</tr>
<tr>
<td><code>term ::=</code></td>
<td>literal constants</td>
</tr>
<tr>
<td>`</td>
<td>id`</td>
</tr>
<tr>
<td>`</td>
<td>unary-op term`</td>
</tr>
<tr>
<td>`</td>
<td>term bin-op term`</td>
</tr>
<tr>
<td>`</td>
<td>term [ term ]`</td>
</tr>
<tr>
<td>`</td>
<td>{ term \with [ term ] = term }`</td>
</tr>
<tr>
<td>`</td>
<td>term . id`</td>
</tr>
<tr>
<td>`</td>
<td>{ term \with . id = term }`</td>
</tr>
<tr>
<td>`</td>
<td>term -&gt; id`</td>
</tr>
<tr>
<td>`</td>
<td>( type-expr ) term`</td>
</tr>
<tr>
<td>`</td>
<td>id ( term (, term)* )`</td>
</tr>
<tr>
<td>`</td>
<td>( term )`</td>
</tr>
<tr>
<td>`</td>
<td>term ? term : term`</td>
</tr>
<tr>
<td>`</td>
<td>\let id = term ; term`</td>
</tr>
<tr>
<td>`</td>
<td>sizeof ( term )`</td>
</tr>
<tr>
<td>`</td>
<td>sizeof ( C-type-name )`</td>
</tr>
<tr>
<td>`</td>
<td>id : term`</td>
</tr>
<tr>
<td>`</td>
<td>string : term`</td>
</tr>
</tbody>
</table>

#### 2.2 Logic expressions

This first section presents the language of expressions one can use in annotations. These are called *logic expressions* in the following. They correspond to pure C expressions, with additional constructs that we will introduce progressively.

Figures 2.1 and 2.2 present the grammar for the basic constructs of logic expressions. In that grammar, we distinguish between *predicates* and *terms*, following the usual distinction between propositions and terms in classical first-order logic. The grammar for binders and type expressions is given separately in Figure 2.3.

With respect to C pure expressions, the additional constructs are as follows:
2.2. LOGIC EXPRESSIONS

Additional connectives C operators & (UTF8: \&) and | (UTF8: |) are used as logical connectives. There are additional connectives => (UTF8: \implies) for implication, <= (UTF8: \iff) for equivalence and ^ (UTF8: ^) for exclusive or. These logical connectives all have a bitwise counterpart, either C ones like &, | and ~, or additional ones like local binding -> and bitwise equivalence <-->.

Quantification Universal quantification is denoted by \forall \tau x_1, ..., x_n; e and existential quantification by \exists \tau x_1, ..., x_n; e.

Local binding \let x = e_1; e_2 introduces the name x for expression e_1 which can be used in expression e_2.

Conditional c ? e_1 : e_2. There is a subtlety here: the condition may be either a boolean term or a predicate. In case of a predicate, the two branches must also be predicates, so that this construct acts as a connective with the following semantics: c ? e_1 : e_2 is equivalent to (c => e_1) & (c => e_2).

Syntactic naming id : e is a term or a predicate equivalent to e. It is different from local naming with \let: the name cannot be reused in other terms or predicates. It is only for readability purposes.

Functional modifier The composite element modifier is an additional operator related to C structure field and array accessors. The expression \{ s .id = v \} denotes the same structure as s, except for the field id that is equal to v. The equivalent expression for an array is \{ t [i] = v \} which returns the same array as t, except for the i\textsuperscript{th} element whose value is v. See section 2.10 for an example of use of these operators.

| rel-op  ::=  == | != | <= | >= | > | < |
| pred   ::=  \true | \false | term (rel-op term)^+ comparisons (see remark) | id (term (, term)* ) predicate application | (pred) parentheses | pred & pred conjunction | pred | pred disjunction | pred => pred implication | pred <=> pred equivalence | ! pred negation | pred ^ pred exclusive or | term ? pred : pred ternary condition | pred ? pred : pred |
| \let id = term ; pred local binding |
| \let id = pred ; pred |
| \forall binders ; pred universal quantification |
| \exists binders ; pred existential quantification |
| id : pred syntactic naming |
| string : pred syntactic naming |

Figure 2.2: Grammar of predicates
CHAPTER 2. SPECIFICATION LANGUAGE

```
binders ::= binder (, binder)*
binder ::= type-expr variable-ident (,variable-ident)*
type-expr ::= logic-type-expr | C-type-name
logic-type-expr ::= built-in-logic-type
                | id
built-in-logic-type ::= boolean | integer | real
variable-ident ::= id | * variable-ident
                | ( variable-ident )
```

Figure 2.3: Grammar of binders and type expressions

**Logic functions** Applications in terms and in propositions are not applications of C functions, but of logic functions or predicates; see Section 2.6 for detail.

**Consecutive comparison operators** The construct \( t_1 \text{ relop}_1 t_2 \text{ relop}_2 t_3 \cdots \text{ relop}_k t_k \) with several consecutive comparison operators is a shortcut for \((t_1 \text{ relop}_1 t_2) \&\& (t_2 \text{ relop}_2 t_3) \&\& \cdots\). It is required that the \( \text{ relop}_i \) operators must be in the same “direction”, i.e. they must all belong either to \{\( <, \leq, =\)\} or to \{\( >, \geq, =\)\}. Expressions such as \( x < y > z \) or \( x != y != z \) are not allowed.

To enforce the same interpretation as in C expressions, one may need to add extra parentheses: \( a == b < c \) is equivalent to \( a == b \&\& b < c \), whereas \( a == (b < c) \) is equivalent to \( \let x = b < c; a == x \). This situation raises some issues, as in the example below.

There is a subtlety regarding comparison operators: they are predicates when used in predicate position, and boolean functions when used in term position.

**Example 2.1** Let us consider the following example:

```
int f(int a, int b) { return a < b; }
```

- the obvious postcondition \( \backslash \text{result} == a < b \) is not the right one because it is actually a shortcut for \( \backslash \text{result} == a \&\& a < b \).
- adding parentheses results in a correct post-condition \( \backslash \text{result} == (a < b) \). Note however that there is an implicit conversion (see Sec. 2.2.3) from the \( \text{int} \) (the type of \( \backslash \text{result} \)) to \( \text{boolean} \) (the type of \( a<b \))
- an equivalent post-condition, which does not rely on implicit conversion, is \( (\backslash \text{result} != 0) == (a<b) \). Both pairs of parentheses are mandatory.
- \( \backslash \text{result} == (\text{integer})(a<b) \) is also acceptable because it compares two integers. The cast towards \( \text{integer} \) enforces \( a<b \) to be understood as a boolean term. Notice that a cast towards \( \text{int} \) would also be acceptable.
- \( \backslash \text{result} != 0 \leftrightarrow a < b \) is acceptable because it is an equivalence between two predicates.
2.2. LOGIC EXPRESSIONS

<table>
<thead>
<tr>
<th>class</th>
<th>associativity</th>
<th>operators</th>
</tr>
</thead>
<tbody>
<tr>
<td>selection</td>
<td>left</td>
<td>[···] --&gt; .</td>
</tr>
<tr>
<td>unary</td>
<td>right</td>
<td>! - + - * &amp; (cast) sizeof</td>
</tr>
<tr>
<td>multiplicative</td>
<td>left</td>
<td>* / %</td>
</tr>
<tr>
<td>additive</td>
<td>left</td>
<td>+ -</td>
</tr>
<tr>
<td>shift</td>
<td>left</td>
<td>&lt;&lt; &gt;&gt;</td>
</tr>
<tr>
<td>comparison</td>
<td>left</td>
<td>&lt; &lt;= &gt; &gt;=</td>
</tr>
<tr>
<td>bitwise and</td>
<td>left</td>
<td>&amp;</td>
</tr>
<tr>
<td>bitwise xor</td>
<td>left</td>
<td>^</td>
</tr>
<tr>
<td>bitwise or</td>
<td>left</td>
<td></td>
</tr>
<tr>
<td>bitwise implies</td>
<td>left</td>
<td>--&gt;</td>
</tr>
<tr>
<td>bitwise equiv</td>
<td>left</td>
<td>&lt;---&gt;</td>
</tr>
<tr>
<td>connective and</td>
<td>left</td>
<td>&amp;&amp;</td>
</tr>
<tr>
<td>connective xor</td>
<td>left</td>
<td>^</td>
</tr>
<tr>
<td>connective or</td>
<td>left</td>
<td></td>
</tr>
<tr>
<td>connective implies</td>
<td>right</td>
<td>==&gt;</td>
</tr>
<tr>
<td>connective equiv</td>
<td>left</td>
<td>&lt;==&gt;</td>
</tr>
<tr>
<td>ternary connective</td>
<td>right</td>
<td>...?...;...</td>
</tr>
<tr>
<td>binding</td>
<td>left</td>
<td>\forall \exists \let</td>
</tr>
<tr>
<td>naming</td>
<td>right</td>
<td>:</td>
</tr>
</tbody>
</table>

Figure 2.4: Operator precedence

2.2.1 Operators precedence

The precedence of C operators is conservatively extended with additional operators, as shown in Figure 2.4. In this table, operators are sorted from highest to lowest priority. Operators of same priority are presented on the same line.

There is a remaining ambiguity between the connective ···?:···::··· and the labelling operator ::. Consider for instance the expression `x?y:z:t`. The precedence table does not indicate whether this should be understood as `x?(y:z):t` or `x?y:(z:t)`. Such a case must be considered as a syntax error, and should be fixed by explicitly adding parentheses.

2.2.2 Semantics

The semantics of logic expressions in ACSL is based on mathematical first-order logic [24]. In particular, it is a 2-valued logic with only total functions. Consequently, expressions are never “undefined”. This is an important design choice and the specification writer should be aware of that. (For a discussion about the issues raised by such design choices, in similar specification languages such as JML, see the comprehensive list compiled by Patrice Chalin [4, 5].)

Having only total functions implies than one can write terms such as `1/0`, or `*p` when `p` is null (or more generally when it points to a non-properly allocated memory cell). In particular, the predicates `1/0 == 1/0` and `*p == *p` are valid, since they are instances of the axiom `∀x, x = x` of first-order logic. The reader should not be alarmed, because there is no way to deduce anything useful from such terms. As usual, it is up to the specification designer to write consistent assertions. For example, when introducing the following lemma (see Section 2.6):
CHAPTER 2. SPECIFICATION LANGUAGE

```plaintext
/*@ lemma div_mul_identity:
forall real x, real y; y != 0.0 ==> y*(x/y) == x; @*/
```

a premise is added to require \( y \) to be non-zero.

2.2.3 Typing

The language of logic expressions is typed (as in multi-sorted first-order logic). Types are either C types or logic types defined as follows:

- “mathematical” types: `integer` for unbounded, mathematical integers, `real` for real numbers, `boolean` for booleans (with values written `true` and `false`);
- logic types introduced by the specification writer (see Section 2.6).

There are implicit coercions for numeric types:

- C integral types `char`, `short`, `int` and `long`, signed or unsigned, are all subtypes of type `integer`;
- `integer` is itself a subtype of type `real`;
- C types `float` and `double` are subtypes of type `real`.

Notes:

- There is a distinction between booleans and predicates. The expression \( x < y \) in term position is a boolean, and the same expression is also allowed in predicate position.
- Unlike in C, there is a distinction between booleans and integers. There is an implicit promotion from integers to booleans, thus one may write \( x \&\& \ y \) instead of \( x != 0 \&\& y != 0 \). If the reverse conversion is needed, an explicit cast is required, *e.g.* `(int)(x>0)+1`, where `false` becomes 0 and `true` becomes 1.
- Quantification can be made over any type: logic types and C types. Quantification over pointers must be used carefully, since it depends on the memory state where dereferencing is done (see Section 2.2.4 and Section 2.6.9).

Formal typing rules for terms are given in appendix A.3.

2.2.4 Integer arithmetic and machine integers

The following integer arithmetic operations apply to mathematical integers: addition, subtraction, multiplication, unary minus. The value of a C variable of an integral type is promoted to a mathematical integer. As a consequence, there is no “arithmetic overflow” in logic expressions.

Division and modulo are also mathematical operations, which coincide with the corresponding C operations on C machine integers, thus following the ISO C99 conventions. In particular, these are not the usual mathematical Euclidean division and remainder. Generally speaking, division rounds the result towards zero. The results are not specified if divisor is zero; otherwise if \( q \) and \( r \) are the quotient and the remainder of \( n \) divided by \( d \) then:
2.2. LOGIC EXPRESSIONS

- \(|d \times q| \leq |n|\), and \(|q|\) is maximal for this property;
- \(q\) is zero if \(|n| < |d|\);
- \(q\) is positive if \(|n| \geq |d|\) and \(n\) and \(d\) have the same sign;
- \(q\) is negative if \(|n| \geq |d|\) and \(n\) and \(d\) have opposite signs;
- \(q \times d + r = n\);
- \(|r| < |d|\);
- \(r\) is zero or has the same sign as \(n\).

Example 2.2 The following examples illustrate the results of division and modulo depending on the sign of their arguments:

- \(5/3\) is 1 and \(5\%3\) is 2;
- \((-5)/3\) is -1 and \((-5)\%3\) is -2;
- \(5/(-3)\) is -1 and \(5\%(-3)\) is 2;
- \((-5)/(-3)\) is 1 and \((-5)\%(-3)\) is -2.

Hexadecimal octal and binary constants

Hexadecimal, octal and binary constants are always non-negative. Suffixes \(u\) and \(l\) for C constants are allowed but meaningless.

Casts and overflows

In logic expressions, casting from mathematical integers to an integral C type \(t\) (such as \(\text{char}\), \(\text{short}\), \(\text{int}\), etc.) is allowed and is interpreted as follows: the result is the unique value of the corresponding type that is congruent to the mathematical result modulo the cardinal of this type, that is \(2^\times \text{sizeof}(t)\).

Example 2.3 \((\text{unsigned char})1000\) is 1000 mod 256 \(i.e.\) 232.

To express in the logic the value of a C expression, one has to add all the necessary casts. For example, the logic expression denoting the value of the C expression \(x*y+z\) is \((\text{int})(\text{int}(x*y)+z)\). Note that there is no implicit cast from integers to C integral types.

Example 2.4 The declaration

```c
//@ logic int f(int x) = x+1 ;
```

is not allowed because \(x+1\), which is a mathematical integer, must be casted to \(\text{int}\). One should write either

```c
//@ logic integer f(int x) = x+1 ;
```

or

```c
//@ logic int f(int x) = (int)(x+1) ;
```
Quantification on C integral types

Quantification over a C integral type corresponds to integer quantification over the corresponding interval.

**Example 2.5** *Thus the formula*

\[
\forall \text{char} \ c; \ c \leq 1000
\]

*is equivalent to*

\[
\forall \text{integer} \ c; \ \text{CHAR_MIN} \leq c \leq \text{CHAR_MAX} \implies c \leq 1000
\]

*where the the bounds **CHAR_MIN** and **CHAR_MAX** are defined in limits.h*

Size of C integer types

The size of C types is architecture-dependent. ACSL does not enforce these sizes either, hence the semantics of terms involving such types is also architecture-dependent. The `sizeof` operator may be used in annotations and is consistent with its C counterpart. For instance, it should be possible to verify the following code:

```c
/*@ ensures \result <= sizeof (int); */
int f() { return sizeof (char); }
```

Constants giving maximum and minimum values of those types may be provided in a library.

Enum types

Enum types are also interpreted as mathematical integers. Casting an integer into an enum in the logic gives the same result as if the cast was performed in the C code.

Bitwise operations

Like arithmetic operations, bitwise operations apply to any mathematical integer: any mathematical integer has a unique infinite 2-complement binary representation with infinitely many zeros (for non-negative numbers) or ones (for negative numbers) on the left. Bitwise operations apply to this representation.

**Example 2.6**

- \( 7 \& 12 = \cdots 00111 \& \cdots 001100 = \cdots 00100 = 4 \)
- \(-8 \mid 5 = \cdots 11000 \mid \cdots 00101 = \cdots 11101 = -3 \)
- \(-5 = \cdots 00101 = \cdots 111010 = -6 \)
- \(-5 << 2 = \cdots 11011 << 2 = \cdots 11101100 = -20 \)
- \( 5 >> 2 = \cdots 00101 >> 2 = \cdots 0001 = 1 \)
- \(-5 >> 2 = \cdots 11011 >> 2 = \cdots 1110 = -2 \)
2.2.5 Real numbers and floating point numbers

Floating-point constants and operations are interpreted as mathematical real numbers: a C variable of type float or double is implicitly promoted to a real. Integers are promoted to reals if necessary. Usual binary operations are interpreted as operators on real numbers, hence they never involve any rounding nor overflow.

**Example 2.7** In an annotation, \(1 \times 10^{300} \times 1 \times 10^{300}\) is equal to \(1 \times 10^{600}\), even if that last number exceeds the largest representable number in double precision: there is no "overflow". 

\(2 \times 0.1\) is equal to the real number \(0.2\), and not to any of its floating-point approximation: there is no "rounding".

Unlike the promotion of C integer types to mathematical integers, there are special float values which do not naturally map to a real number, namely the IEEE-754 special values for “not-a-number”, \(+\infty\) and \(-\infty\). See below for a detailed discussion on such special values. However, remember that ACSL’s logic has only total functions. Thus, there are implicit promotion functions `real_of_float` and `real_of_double` whose results on the 3 values above is left unspecified.

In logic, real literals can also be expressed under the hexadecimal form of C99: \(0x\text{hh}.\text{hhp}\pm dd\) where \(h\) are hexadecimal digits and \(dd\) is in decimal, denotes number \(\text{hh} . \text{hh} \times 2^{dd}\), e.g. \(0x1.Fp-4\) is \((1 + 15/16) \times 2^{-4}\).

Usual operators for comparison are interpreted as real operators too. In particular, equality operation \(\equiv\) of float (or double) expressions means equality of the real numbers they represent respectively. Or equivalently, \(x \equiv y\) for \(x,y\) two float variables means \(\text{real_of_float}(x) \equiv \text{real_of_float}(y)\) with the mathematical equality of real numbers.

Special predicates are also available to express the comparison operators of float (resp. double) numbers as in C: \(\text{eq_float}\), \(\text{gt_float}\), \(\text{ge_float}\), \(\text{le_float}\), \(\text{lt_float}\), \(\text{ne_float}\) (resp. for double).

**Casts, infinity and NaNs**

Casting from a C integer type or a float type to a float or a double is as in C: the same conversion operations apply.

Conversion of real numbers to float or double values indeed depends on various possible rounding modes defined by the IEEE 754 standard [23, 25]. These modes are defined by a logic type (see section 2.6.8):

```c
/*@
 type rounding_mode = Up | Down | ToZero | NearestAway | NearestEven;
 */
```

Then rounding a real number can be done explicitly using functions

```c
logic float round_float(rounding_mode m, real x);
logic double round_double(rounding_mode m, real x);
```

Cast operators (`float`) and `double` applied to a mathematical integer or real number \(x\) are equivalent to apply rounding functions above with the nearest-even rounding mode (which is the default rounding mode in C programs) If the source real number is too large, this may also result into one of the special values `+infinity` and `-infinity`. 
Example 2.8 We have \((\text{float})0.1 \equiv 13421773 \times 2^{-27}\) which is equal to 0.100000001490116119384765625.

Notice also that unlike for integers, suffixes \(f\) and \(l\) are meaningful, because they implicitly add a cast operator as above.

This semantics of casts ensures that the float result \(r\) of a C operation \(e_1 \ op \ e_2\) on floats, if there is no overflow and if the default rounding mode is not changed in the program, has the same real value as the logic expression \((\text{float}) (e_1 \ op \ e_2)\). Notice that this is not true for the equality \(\eq\) of floats: \(-0.0 + -0.0\) in C is equal to the float number \(-0.0\), which is not \(\eq\) to \(0.0\), which is the value of the logic expression \((\text{float})(-0.0 + -0.0)\).

Finally, additional predicates are provided on float and double numbers, which check that their argument is NaN or a finite number respectively:

| 1 | predicate \is_NaN(float x); |
| 2 | predicate \is_NaN(double x); |
| 3 | predicate \is_finite (float x); |
| 4 | predicate \is_finite (double x); |

Quantification

Quantification over a variable of type \text{real} is of course usual quantification over real numbers. Quantification over float (resp. double) types is allowed too, and is supposed to range over all real numbers representable as floats (resp doubles). In particular, this does not include NaN, +infinity and -infinity in the considered range.

Mathematical functions

Classical mathematical operations like exponential, sine, cosine, and such are available as built-in:

| 1 | integer \min(integer x, integer y); |
| 2 | integer \max(integer x, integer y); |
| 3 | real \min(real x, real y); |
| 4 | real \max(real x, real y); |
| 5 | integer \abs(integer x); |
| 6 | real \abs(real x); |
| 7 | real \sqrt(real x); |
| 8 | real \pow(real x, real y); |
| 9 | integer \ceil(real x); |
| 10 | integer \floor(real x); |
| 11 | real \exp(real x); |
| 12 | real \log(real x); |
| 13 | real \log10(real x); |
| 14 | real \cos(real x); |
| 15 | real \sin(real x); |
| 16 | real \tan(real x); |
| 17 | real \cosh(real x); |
| 18 | real \sinh(real x); |
2.2 LOGIC EXPRESSIONS

\[
\begin{align*}
\text{real } \tanh(\text{real } x) ; \\
\text{real } \acos(\text{real } x) ; \\
\text{real } \asin(\text{real } x) ; \\
\text{real } \atan(\text{real } x) ; \\
\text{real } \atan2(\text{real } y, \text{real } x) ; \\
\text{real } \hypot(\text{real } x, \text{real } y) ;
\end{align*}
\]

Exact computations

In order to specify properties of rounding errors, it is useful to express something about the so-called exact computations [3]: the computations that would be performed in an ideal mode where variables denote true real numbers.

To express such exact computations, two special constructs exist in annotations:

- \( \texttt{\textbackslash exact}(x) \) denotes the value of the C variable \( x \) (or more generally any C left-value) as if the program was executed with ideal real numbers.
- \( \texttt{\textbackslash round\_error}(x) \) is a shortcut for \( |x - \texttt{\textbackslash exact}(x)| \)

Example 2.9 Here is an example of a naive approximation of cosine [2].

```c
/*@ requires \texttt{\textbackslash abs}(\texttt{\textbackslash exact}(x)) <= 0x1p-5; 
@ requires \texttt{\textbackslash round\_error}(x) <= 0x1p-20; 
@ ensures \texttt{\textbackslash abs}(\texttt{\textbackslash exact}(\texttt{\textbackslash result}) - \texttt{\textbackslash cos}(\texttt{\textbackslash exact}(x))) <= 0x1p-24; 
@ ensures \texttt{\textbackslash round\_error}(\texttt{\textbackslash result}) <= \texttt{\textbackslash round\_error}(x) + 0x3p-24; 
@*/
float cosine(float x) {
    return 1.0f - x * x * 0.5f;
}
```

2.2.6 C arrays and pointers

Address operator, array access, pointer arithmetic and dereferencing

These operators are similar to their corresponding C operators.

\texttt{\textbackslash address-of} operator should be used with caution. Values in logic do not lie in C memory so it does not mean anything to talk about their “address”.

Unlike in C, there is no implicit cast from an array type to a pointer type. Nevertheless, arithmetic and dereferencing over arrays lying in C memory are allowed like in C.

Example 2.10 Dereferencing a C array is equivalent to an access to the first element of the array; shifting it from \( i \) denotes the address of its \( i^{\text{th}} \) element.

```c
int tab[10] = { 1 }; 
int x; 
int *p = &x; 
//@ requires p == &x
int main(void) {
    //@ assert tab[0]==1 && *p == x;
```
Since pointers can only refer to values lying in C memory, \( p->s \) is always equivalent to \((*p).s\). On the contrary, \( t[i] \) is not always equivalent to \(*({t+i})\), especially for arrays not lying in C memory. Section 2.2.7 details the use of arrays as logic values. There are also differences between \( t \) and the pointer to its first element when evaluating an expression at a given program point. See Section 2.4.3 for more information.

**Function pointers**

Pointers to C functions are allowed in logic. The only possible use of them is to check for equality.

**Example 2.11**

```c
int f(int x);
int g(int x);
//@ requires p == &f || p == &g;
void h(int(*p)(int)) {
...
}
```

### 2.2.7 Structures, Unions and Arrays in logic

Aggregate C objects (i.e. structures, unions and arrays) are also possible values for terms in logic. They can be passed as parameters (and also returned) to logic functions, tested for equality, etc. like any other values.

Aggregate types can be declared in logic, and their contents may be any logic types themselves. Constructing such values in logic can be performed using a syntax similar to C designated initializers.

**Example 2.12** Array types in logic may be declared either with or without an explicit non-negative length. Access to the length of a logic array can be done with \( \text{\textbackslash length} \).
Beware that because of the principle of only total functions in logic, \( t[i] \) can appear in ACSL annotations even if \( i \) is outside the array bounds.

**Functional updates**

Syntax for functional update is similar to initialization of aggregate objects.

**Example 2.13** Functional update of an array is done by

\[
\{ \ t_{iso} \with [0] = \{ .x = 3.0, .y = 3.0 \} \}
\]

Functional update of a structure is done by

\[
\{ \ \operatorname{origin} \with .x = 3.0 \}
\]

There is no particular syntax for functional update of an union. For an object of an union type, the following equality is not true

\[
\{ \ { \ \operatorname{object} \with .x = 3.0 \} \with .y = 2.0 \} == \{ \ { \ \operatorname{object} \with .y = 2.0 \} \with .x = 3.0 \}
\]

The equality predicate \( == \) applies to aggregate values, but it is required that they have the same type. Then equality amounts the recursively check equality of fields. Equality of arrays of different lengths returns false. Beware that equality of unions is also equality of all fields.

**C aggregate types**

C aggregates types (struct, union or array) naturally map to logic types, by recursively mapping their fields.

**Example 2.14** There is no implicit cast to type of the updated/initialized fields.

\[
\text{struct } S \{ \ \text{int } x; \ \text{float } y; \ \text{int } t[10]; \}\;
\]

//@ logic integer f(struct S s) = s.t[3];
//@ logic struct S g(integer n, struct S s) = { s \with .x = (\text{int})n };
Cast and conversion

Unlike in C, there is no implicit conversion from an array type to a pointer type. On the other hand, there is an implicit conversion from an array of a given size to an array with unspecified size (but not the converse).

Example 2.15

```c
//@ logic point square[4] = { origin, ... };
//@ ... perimeter(square);    // well typed
//@ ... centroid(square);     // wrongly typed
//@ ... centroid((triangle)square); // well-typed (truncation)
```

An explicit cast from an array type to a pointer type is allowed only for arrays that lies in C memory. As in C, the result of the cast is the address of the first element of the array (see Section 2.2.6).

Conversely, an explicit cast from a pointer type to an array type or a structure type is allowed, and acts as collecting the values it points to.

Subtyping and cast recursively applies to fields.

Example 2.16

```c
struct { float u,v; } p[10];
//@ assert centroid((point[3])p) == ...
//@ assert perimeter((point[])p) == ...
```

Precisely, conversion of a pointer `p` of type `τ*` to an logic array of type `τ[]` returns a logic array `t` such that

\[
\text{length}(t) = (\text{\textbackslash block\_length}(p) - \text{\textbackslash offset}(p)) / \text{sizeof}(τ)
\]

2.2.8 String literals

2.3 Function contracts

Figure 2.5 shows a grammar for function contracts. Location denotes a memory location and is defined in Section 2.3.4. Allocation-clauses allow to specify which memory location is dynamically allocated or deallocated by the function from the heap; they are defined later in Section 2.7.3.

This section is organized as follows. First, the grammar for terms is extended with two new constructs. Then Section 2.3.2 introduces simple contracts. Finally, Section 2.3.3 defines more general contracts involving named behaviors.

The decreases and terminates clauses are presented later in Section 2.5. Abrupt-clauses allow to specify what happens when the function does not return normally but exits abruptly; they are defined in Section 2.9.
2.3 FUNCTION CONTRACTS

function-contract ::= requires-clause* terminates-clause? decreases-clause? simple-clause* named-behavior* completeness-clause*
  
  requires-clause ::= requires pred ;
  terminates-clause ::= terminates pred ;
  decreases-clause ::= decreases term (for id)? ;
  simple-clause ::= assigns-clause | ensures-clause | allocation-clause | abrupt-clause-fn
  assigns-clause ::= assigns locations ;
  locations ::= location (, location)* | \nothing
  location ::= tset
  ensures-clause ::= ensures pred ;
  named-behavior ::= behavior id : behavior-body
  behavior-body ::= assumes-clause* requires-clause* simple-clause*
  assumes-clause ::= assumes pred ;
  completeness-clause ::= complete behaviors (id (, id)*)? ;
  | disjoint behaviors (id (, id)*)? ;

Figure 2.5: Grammar of function contracts

term ::= \old ( term ) old value | \result result of a function

pred ::= \old ( pred )

Figure 2.6: \old and \result in terms

2.3.1 Built-in constructs \old and \result

Post-conditions usually require to refer to both the function result and values in the pre-state. Thus terms are extended with the following new constructs (shown in Figure 2.6).

- \old (e) denotes the value of predicate or term e in the pre-state of the function.

- \result denotes the returned value of the function.

\old (e) and \result can be used only in ensures clauses, since the other clauses already refer to the pre-state. In addition, \result can not be used in the contract of a function which returns void.

C function parameters are obtained by value from actual parameters that mostly remain unaltered by the function calls. For that reason, formal parameters in function contracts are defined such that they always refer implicitly to their values interpreted in the pre-state. Thus, \old construct is useless for formal parameters (in function contracts only).
2.3.2 Simple function contracts

A simple function contract, having only simple clauses and no named behavior, takes the following form:

```plaintext
/*@ requires P_1; requires P_2; ...
@ assigns L_1; assigns L_2; ...
@ ensures E_1; ensures E_2; ...
@*/
```

The semantics of such a contract is as follows:

- The caller of the function must guarantee that it is called in a state where the property $P_1 \&\& P_2 \&\& \ldots$ holds.
- The called function returns a state where the property $E_1 \&\& E_2 \&\& \ldots$ holds.
- All memory locations that are allocated in both the pre-state and the post-state do not belong to the set $L_1 \cup L_2 \cup \ldots$ are left unchanged in the post-state. The set $L_1 \cup L_2 \cup \ldots$ itself is interpreted in the pre-state.

Having multiple `requires`, `assigns`, or `ensures` clauses only improves readability since the contract above is equivalent to the following simplified one:

```plaintext
/*@ requires P_1 \&\& P_2 \&\& \ldots;
@ assigns L_1, L_2, \ldots;
@ ensures E_1 \&\& E_2 \&\& \ldots;
@*/
```

If no clause `requires` is given, it defaults to `\true`, and similarly for `ensures` clause. Giving no `assigns` clause means that locations assigned by the function are not specified, so the caller has no information at all on this function’s side effects. See Section 2.3.5 for more details on default status of clauses.

Example 2.17 The following function is given a simple contract for computation of the integer square root.

```plaintext
/*@ requires x >= 0;
@ ensures \result >= 0;
@ ensures \result \ast \result <= x;
@ ensures x < (\result + 1) \ast (\result + 1);
@*/
int isqrt(int x);
```

The contract means that the function must be called with a nonnegative argument, and returns a value satisfying the conjunction of the three `ensures` clauses. Inside these `ensures` clauses, the use of the construct `\old(x)` is not necessary, even if the function modifies the formal parameter `x`, because function calls modify a copy of the effective parameters, and the effective parameters remain unaltered. In fact, `x` denotes the effective parameter of `isqrt` calls which has the same value interpreted in the pre-state than in the post-state.

Example 2.18 The following function is given a contract to specify that it increments the value pointed to by the pointer given as argument.

---

1Functions that allocate or free memory can be specified with additional clauses described in section 2.7.3.
The contract means that the function must be called with a pointer \( p \) that points to a safely allocated memory location (see Section 2.7 for details on the \( \text{\texttt{\textbackslash valid}} \) built-in predicate). It does not modify any memory location but the one pointed to by \( p \). Finally, the \texttt{ensures} clause specifies that the value \( *p \) is incremented by one.

### 2.3.3 Contracts with named behaviors

The general form of a function contract contains several named behaviors (restricted to two behaviors, in the following, for readability).

The semantics of such a contract is as follows:

- The caller of the function must guarantee that the call is performed in a state where the property \( P \land (A_1 \implies R_1) \land (A_2 \implies R_2) \) holds.

- The called function returns a state where the properties \( \text{\texttt{\textbackslash old}}(A_i) \implies E_i \) hold for each \( i \).

- For each \( i \), if the function is called in a pre-state where \( A_i \) holds, then each memory location of that pre-state that does not belong to the set \( L_i \) is left unchanged in the post-state.

\texttt{requires} clauses in the behaviors are proposed mainly to improve readability (to avoid some duplication of formulas), since the contract above is equivalent to the following simplified one:

```c
/*\* requires \ valid (p);
  \* assigns *p;
  \* ensures *p == \old(*p) + 1;
  @*/
void incrstar(int *p);
```
A simple contract such as

```plaintext
/*@ requires P; assigns L; ensures E; */
```

is actually equivalent to a single named behavior as follows:

```plaintext
/*@ requires P;
@ behavior <any name>:
@ assumes \true;
@ assigns L;
@ ensures E;
@*/
```

Similarly, global assigns and ensures clauses are equivalent to a single named behavior. More precisely, the following contract

```plaintext
/*@ requires P;
@ assigns L;
@ ensures E;
@ behavior b_1: ... 
@ behavior b_2: ...
@ ... 
@*/
```

is equivalent to

```plaintext
/*@ requires P;
@ behavior <any name>:
@ assumes \true;
@ assigns L;
@ ensures E;
@ behavior b_1: ...
@ behavior b_2: ...
@ ... 
@*/
```

**Example 2.19** In the following, `bsearch(t,n,v)` searches for element `v` in array `t` between indices 0 and `n-1`.

```plaintext
/*@ requires n >= 0 && \valid(t+(0..n-1));
@ assigns \nothing;
@ ensures -1 <= \result <= n-1;
@ behavior success:
@ ensures \result >= 0 ==> t[\result] == v;
@ behavior failure:
@ assumes t_is_sorted : \forall integer k1, integer k2;
@ 0 <= k1 <= k2 <= n-1 ==> t[k1] <= t[k2];
@ ensures \result == -1 ==> \forall integer k; 0 <= k < n ==> t[k] != v;
@*/
int bsearch(double t[], int n, double v);
```

The precondition requires array `t` to be allocated at least from indices 0 to `n-1`. The two named behaviors correspond respectively to the successful behavior and the failing behavior.

Since the function is performing a binary search, it requires the array `t` to be sorted in increasing order: this is the purpose of the predicate named `t_is_sorted` in the assumes clause of the behavior named failure.

See 2.4.2 for a continuation of this example.
Example 2.20 The following function illustrates the importance of different assigns clauses for each behavior:

```c
/*@ behavior p_changed:
  @ assumes n > 0;
  @ requires \valid(p);
  @ assigns *p;
  @ ensures *p == n;
*/

void f(int n, int *p, int *q) {
  if (n > 0) *p = n;
  else *q = n;
}

/*@ behavior q_changed:
  @ assumes n <= 0;
  @ requires \valid(q);
  @ assigns *q;
  @ ensures *q == n;
*/
```

Its contract means that it assigns values pointed to by p or by q, conditionally on the sign of n.

Completeness of behaviors

In a contract with named behaviors, it is not required that the disjunction of the \( A_i \) is true, \( i.e. \) it is not mandatory to provide a “complete” set of behaviors. If such a condition is sought, it is possible to add the following clause to a contract:

```c
/*@ ... @/  
  @ complete behaviors b_1, ..., b_n;  
  @*/
```

It specifies that the set of behaviors \( b_1, ..., b_n \) is complete \( i.e. \) that

\[
R \implies (A_1 \lor A_2 \lor ... \lor A_n)
\]

holds, where \( R \) is the precondition of the contract. The simplified version of that clause

```c
/*@ ... @/  
  @ complete behaviors;  
  @*/
```

means that all behaviors given in the contract should be taken into account.

Similarly, it is not required that two distinct behaviors are disjoint. If desired, this can be specified with the following clause:

```c
/*@ ... @/  
  @ disjoint behaviors b_1, ..., b_n;  
  @*/
```

It means that the given behaviors are pairwise disjoint \( i.e. \) that, for all distinct \( i \) and \( j \),

\[
R \implies \neg (A_i \land A_j)
\]

holds. The simplified version of that clause

```c
/*@ ... @/  
  @ disjoint behaviors;  
  @*/
```

means that all behaviors given in the contract should be taken into account. Multiple complete and disjoint sets of behaviors can be given for the same contract.
2.3.4 Memory locations and sets of terms

There are several places where one needs to describe a set of memory locations: in assigns clauses of function contracts, or in loop assigns clauses (see section 2.4.2). A memory location is then any set of terms denoting a set of l-values. Moreover, a location given as argument to an assigns clause must be a set of modifiable l-values, as described in Section A.1. More generally, we introduce syntactic constructs to denote sets of terms which are also useful for \separated predicate (see Section 2.7.2)

The grammar for sets of terms is given in Figure 2.7. The semantics is given below, where \( s \) denotes any \( \mathrm{tset} \).

- \( \emptyset \) denotes the empty set.
- a simple term denotes a singleton set.
- \( s \rightarrow id \) denotes the set of \( x \rightarrow id \) for each \( x \in s \).
- \( s \cdot id \) denotes the set of \( x \cdot id \) for each \( x \in s \).
- \( \ast s \) denotes the set of \( \ast x \) for each \( x \in s \).
- \( \& s \) denotes the set of \( \& x \) for each \( x \in s \).
- \( s_1[s_2] \) denotes the set of \( x_1[x_2] \) for each \( x_1 \in s_1 \) and \( x_2 \in s_2 \).
- \( t_1 \ldots t_2 \) denotes the set of integers between \( t_1 \) and \( t_2 \), included. If \( t_1 > t_2 \), this the same as \( \emptyset \)
- \( \cup \{ \, tset \mid binders (; pred)^7 \} \) set comprehension
- \( \{ \, term \} \) explicit singleton
- \( \{ \, term \} \) implicit singleton
- \( \subset ( \, tset , tset \, ) \) set inclusion

\[ \begin{align*}
\text{tset} & ::= \emptyset \quad \text{empty set} \\
& | \text{tset} \rightarrow \text{id} \\
& | \text{tset} \cdot \text{id} \\
& | \ast \text{tset} \\
& | \& \text{tset} \\
& | \text{tset} \lbrack \text{tset} \rbrack \\
& | \text{term}^\ldots \cdot \text{term}^? \quad \text{range} \\
& | \cup ( \, \text{tset} , \text{tset}^\ast \, ) \quad \text{union of locations} \\
& | \inter ( \, \text{tset} , \text{tset}^\ast \, ) \quad \text{intersection} \\
& | \text{tset} + \text{tset} \\
& | ( \, \text{tset} \, ) \\
& | \{ \, \text{tset} \mid \text{binders} ; \text{pred}^7 \} \quad \text{set comprehension} \\
& | \{ \, \text{term} \} \quad \text{explicit singleton} \\
& | \text{term} \quad \text{implicit singleton} \\
\end{align*} \]

\( \text{pred} ::= \subset ( \, \text{tset} , \text{tset} \, ) \) set inclusion

Figure 2.7: Grammar for sets of terms
2.3. FUNCTION CONTRACTS

- \( \langle s \rangle \) denotes the same set as \( s \);
- \( \{ s \mid b ; P \} \) denotes set comprehension, that is the union of the sets denoted by \( s \) for each value \( b \) of binders satisfying predicate \( P \) (binders \( b \) are bound in both \( s \) and \( P \)).

Note that \( \text{assigns} \ \\emptyset \) is equivalent to \( \text{assigns} \ \\emptyset \); it is left for convenience.

**Example 2.21** The following function sets to 0 each cell of an array.

```c
/*@ requires \valid (t+(0..n-1));
@ assigns t[0..n-1];
@ assigns *(t+(0..n-1));
@ assigns *(t+i i integer i; 0 <= i < n));
@*/
void reset_array(int t[],int n) {
    int i;
    for (i=0; i < n; i++) t[i] = 0;
}
```

*It is annotated with three equivalent \text{assigns} clauses, each one specifying that only the set of cells \( \{t[0],...,t[n-1]\} \) is modified.*

**Example 2.22** The following function increments each value stored in a linked list.

```c
struct list {
    int hd;
    struct list *next;
};

//@ inductive reachable{L}(struct list *root, struct list *to) {
@ case empty{L}: \forall struct list *l; reachable(l,l) ;
@ case non_empty{L}:
    \forall struct list *l1,*l2;
    @ valid(l1) && reachable(l1->next,l2) ==> reachable(l1,l2) ;
@ }
@*/

void incr_list(struct list *p) {
    while (p) { p->hd++ ; p = p->next; }
}
```

*The \text{assigns} clause specifies that the set of modified memory locations is the set of fields \( q->hd \) for each pointer \( q \) reachable from \( p \) following \text{next} fields. See Section 2.6.3 for details about the declaration of the predicate \text{reachable}.*

2.3.5 Default contracts, multiple contracts

A C function can be defined only once but declared several times. It is allowed to annotate each of these declarations with contracts. Those contracts are seen as a single contract with the union of the \text{requires} clauses and behaviors.
On the other hand, a function may have no contract at all, or a contract with missing clauses. Missing requires and ensures clauses default to `true`. If no assigns clause is given, it remains unspecified. If the function under consideration has only a declaration but no body, then it means that it potentially modifies “everything”, hence in practice it will be impossible to verify anything about programs calling that function; in other words giving it a contract is in practice mandatory. On the other hand, if that function has a body, giving no assigns clause means in practice that it is left to tools to compute an over-approximation of the sets of assigned locations.

### 2.4 Statement annotations

Annotations on C statements are of three kinds:

- **Assertions**: allowed before any C statement or at end of blocks.
- **Loop annotations**: invariant, assigns clause, variant; allowed before any loop statement: while, for, and do ... while.
- **Statement contracts**: allowed before any C statement, specifying their behavior in a similar manner to C function contracts.

#### 2.4.1 Assertions

The syntax of assertions is given in Figure 2.8, as an extension of the grammar of C statements.

- `assert` \( P \) means that \( P \) must hold in the current state (the sequence point where the assertion occurs).
- The variant for \( \text{id}_1, \ldots, \text{id}_k : \text{assert} \ P \) associates the assertion to the named behaviors \( \text{id}_i \), each of them being a behavior identifier for the current function (or a behavior of an enclosing block as defined later in Section 2.4.4). It means that this assertion must hold only for the considered behaviors.

#### 2.4.2 Loop annotations

The syntax of loop annotations is given in Figure 2.9, as an extension of the grammar of C statements. Loop-allocation clauses allow to specify which memory location is dynamically allocated or deallocated by a loop from the heap; they are defined later in Section 2.7.3.
2.4. STATEMENT ANNOTATIONS

![Grammar for loop annotations]

Loop invariants and loop assigns

The semantics of loop invariants and loop assigns is defined as follows: a simple loop annotation of the form

```plaintext
/*@ loop-invariant I; @*/
```

specifies that the following conditions hold.

- The predicate \( I \) holds before entering the loop (in the case of a \( \text{for} \) loop, this means right after the initialization expression).
- The predicate \( I \) is an inductive invariant, that is if \( I \) is assumed true in some state where the condition \( c \) is also true, and if execution of the loop body in that state ends normally at the end of the body or with a \( \text{continue} \) statement, \( I \) is true in the resulting state. If the loop condition has side effects, these are included in the loop body in a suitable way:
  - for a \( \text{while} (c) \) loop, \( I \) must be preserved by the side-effects of \( c \) followed by \( s \);
  - for a \( \text{for} (\text{init};c;\text{step}) \) loop, \( I \) must be preserved by the side-effects of \( c \) followed by \( s \) followed by \( s \);
  - for a \( \text{do s while (c)} \) loop, \( I \) must be preserved by \( s \) followed by the side-effects of \( c \).

Note that if \( c \) has side-effects, the invariant might not be true at the exit of the loop: the last “step” starts from a state where \( I \) holds, performs the side-effects of \( c \), which
in the end evaluates to false and exits the loop. Likewise, if a loop is exited through a
\textbf{break} statement, \(I\) does not necessarily hold, as side effects may occur between the last
state in which \(I\) was supposed to hold and the \textbf{break} statement.

- At any loop iteration, any location that was allocated before entering the loop, and is
not member of \(L\) (interpreted in the current state) has the same value as before entering
the loop. In fact, the \textbf{loop assigns} clause specifies an inductive invariant for the locations
that are not member of \(L\).

\textbf{Loop behaviors}

A loop annotation preceded by for \textit{id}_1,\ldots,\textit{id}_k: is similar as above, but applies only for
behaviors \textit{id}_1,\ldots,\textit{id}_k of the current function, hence in particular holds only under the
assumption of their assumes clauses.

\textbf{Remarks}

- The \texttt{old} construct is not allowed in loop annotations. The \texttt{at} form should be used to
refer to another state (see Section 2.4.3).

- When a loop exits with \textbf{break} or \textbf{return} or \textbf{goto}, it is not required that the loop invariant
holds. In such cases, locations that are not member of \(L\) can be assigned between the
end of the previous iteration and the exit statement.

\textbf{Example 2.23} Here is a continuation of example 2.19. Note the use of a loop invariant
associated to a function behavior.

```c
int bsearch(double t[], int n, double v) {
    int l = 0, u = n-1;
    /*@ loop invariant 0 <= 1 && u <= n-1;
    @ for failure: loop invariant
    @   forall integer k; 0 <= k < n && t[k] == v ==> 1 <= k <= u;
    */
    while (1 <= u ) {
        int m = l + (u-l)/2; // better than (l+u)/2
        if (t[m] < v) l = m + 1;
        else if (t[m] > v) u = m - 1;
        else return m;
    }
    return -1;
}
```
2.4. STATEMENT ANNOTATIONS

| assertion ::= /*@ invariant pred ; */ |
|             | /*@ for id (, id)* : invariant pred ; */ |

Figure 2.10: Grammar for general inductive invariants

Loop variants

Optionally, a loop annotation may include a loop variant of the form

| /*@ loop variant m; */ |

where m is a term of type integer.

The semantics is as follows: for each loop iteration that terminates normally or with continue, the value of m at end of the iteration must be smaller that its value at the beginning of the iteration. Moreover, its value at the beginning of the iteration must be nonnegative. Note that the value of m at loop exit might be negative. It does not compromise termination of the loop. Here is an example:

Example 2.24

| void f(int x) { |
| //@ loop variant x; |
| while (x >= 0) { |
| x -= 2; |
| } |
| } |

It is also possible to specify termination orderings other than the usual order on integers, using the additional for modifier. This is explained in Section 2.5.

General inductive invariants

It is actually allowed to pose an inductive invariant anywhere inside a loop body. For example, it makes sense for a do s while (c); loop to contain an invariant right after statement s. Such an invariant is a kind of assertion, as shown in Figure 2.10.

Example 2.25 In the following example, the natural invariant holds at this point (max and lambda are introduced later in Section 2.6.7). It would be less convenient to set an invariant at the beginning of the loop.

| /*@ requires n > 0 &\& \valid(t+(0..n-1)); |
| \@ ensures \result == \max(0,n-1,\lambda integer k ; t[k])); |
| @*/ |
| double max(double t[], int n) { |
| int i = 0; double m,v; |
| do { |
| v = t[i++]; |
| m = v > m ? v : m; |
| /*@ invariant m == \max(0,i-1,\lambda integer k ; t[k])); */ |
| } while (i < n); |
| return m; |
| } |
More generally, loops can be introduced by \texttt{goto}s. As a consequence, such invariants may occur anywhere inside a function’s body. The meaning is that the invariant holds at that point, much like an \texttt{assert}. Moreover, the invariant must be inductive, \textit{i.e.} it must be preserved across a loop iteration. Several invariants are allowed at different places in a loop body. These extensions are useful when dealing with complex control flows.

\textbf{Example 2.26} Here is a program annotated with an invariant inside the loop body:

```c
/*@ requires n > 0; */
@ ensures \result == \max(0,n-1,\lambda integer k; t[k]);
@*/
double max_array(double t[], int n) {
    double m; int i=0;
goto L;
do {
    if (t[i] > m) { L: m = t[i]; }
    /*@
invariant
    @ 0 <= i < n && m == \max(0,i,\lambda integer k; t[k]);
    @*/
    i++;
} while (i < n);
return m;
}
```

The control-flow graph of the code is as follows

```
\begin{tikzpicture}
    \node[circle,fill] (i0) at (0,0) {$0$};
    \node[circle,fill] (m0) at (1,1) {$m < t[i]$};
    \node[circle,fill] (i1) at (2,2) {$i < n$};
    \node[circle,fill] (m1) at (3,3) {$m \geq t[i]$};
    \node[circle,fill] (i2) at (4,4) {$i \geq n$};
    \node[circle,fill] (inv) at (2,6) {\texttt{inv}};
    \node[circle,fill] (do) at (2,8) {\texttt{do}};
    \node[circle,fill] (l) at (2,10) {$\texttt{L}$};

    \draw[->] (i0) -- (m0);
    \draw[->] (m0) -- (i1);
    \draw[->] (i1) -- (m1);
    \draw[->] (m1) -- (i2);
    \draw[->] (i0) -- (l);
    \draw[->] (l) -- (do);
    \draw[->] (do) -- (inv);
    \draw[->] (inv) -- (m0);
    \draw[->] (inv) -- (i1);
    \draw[->] (i2) -- (inv);

    \node at (0.5,0.5) {$m < t[i]$};
    \node at (1.5,1.5) {$m \geq t[i]$};
    \node at (2.5,2.5) {$i < n$};
    \node at (3.5,3.5) {$i = i + 1$};
    \node at (4.5,4.5) {$i \geq n$};
    \node at (2,7) {\texttt{inv}};
    \node at (2,9) {\texttt{do}};
    \node at (2,11) {$\texttt{L}$};
\end{tikzpicture}
```

The invariant is inductively preserved by the two paths that go from node “\texttt{inv}” to itself.

\textbf{Example 2.27} The program

```c
int x = 0;
int y = 10;
/*@ loop invariant 0 <= x < 11; */
while (y > 0) {
    x++;
y--;
}
```
is not correctly annotated, even if it is true that \( x \) remains smaller than 11 during the execution. This is because it is not true that the property \( x<11 \) is preserved by the execution of \( x++ ; y--; \). A correct loop invariant could be \( 0 <= x < 11 \&\& x+y == 10 \). It holds at loop entrance and is preserved (under the assumption of the loop condition \( y>0 \)).

Similarly, the following general invariants are not inductive:

```plaintext
1  int x = 0;
2  int y = 10;
3  while (y > 0) {
4      x++;
5      //@ invariant 0 < x < 11;
6      y--;
7      //@ invariant 0 <= y < 10;
8  }
```

since \( 0 <= y < 10 \) is not a consequence of hypothesis \( 0 < x < 11 \) after executing \( y--; \) and \( 0 < x < 11 \) cannot be deduced from \( 0 <= y < 10 \) after looping back through the condition \( y>0 \) and executing \( x++ \). Correct invariants could be:

```plaintext
1  while (y > 0) {
2      x++;
3      //@ invariant 0 < x < 11 \&\& x+y == 11;
4      y--;
5      //@ invariant 0 <= y < 10 \&\& x+y == 10;
6  }
```

### 2.4.3 Built-in construct `\at`

Statement annotations usually need another additional construct `\at(e, id)` referring to the value of the expression \( e \) in the state at label \( id \). In particular, for a C array of int, \( t \), `\at(t, id)` is a logical array whose content is the same as the one of \( t \) in state at label \( id \). It is thus very different from `\at((int *)t, id)`, which is a pointer to the first element of \( t \) (and stays the same between the state at \( id \) and the current state). Namely, if \( t[0] \) has changed since \( id \), we have `\at(t, id)[0] \neq \at((int *)t, id)[0]`.

The label \( id \) can be either a regular C label, or a label added within a ghost statement as described in Section 2.12. This label must be declared in the same function as the occurrence of `\at(e, id)`, but unlike gotos, more restrictive scoping rules must be respected:

- the label \( id \) must occur before the occurrence of `\at(e, id)` in the source;
- the label \( id \) must not be inside an inner block.

These rules are exactly the same rules as for the visibility of local variables within C statements (see [13], Section A11.1).

### Default logic labels

There are six predefined logic labels: Pre, Here, Old, Post, LoopEntry and LoopCurrent. `\old(e)` is in fact syntactic sugar for `\at(e, Old)`.
• The label \texttt{Here} is visible in all statement annotations, where it refers to the state where the annotation appears; and in all contracts, where it refers to the pre-state for the \texttt{requires}, \texttt{assumes}, \texttt{assigns}, \texttt{variant}, \texttt{terminates}, clauses and the post-state for other clauses. It is also visible in data invariants, presented in Section 2.11.

• The label \texttt{Old} is visible in \texttt{assigns} and \texttt{ensures} clauses of all contracts (both for functions and for statement contracts described below in Section 2.4.4), and refers to the pre-state of this contract.

• The label \texttt{Pre} is visible in all statement annotations, and refers to the pre-state of the function it occurs in.

• The label \texttt{Post} is visible in \texttt{assigns} and \texttt{ensures} clauses of all contracts, and it refers to the post-state.

• The label \texttt{LoopEntry} is visible in loop annotations and all annotations related to a statement enclosed in a loop. It refers to the state just before entering that loop for the first time—but after initialization took place in the case of a \texttt{for} loop, as for \texttt{loop invariant} (section 2.4.2). When \texttt{LoopEntry} is used in a statement enclosed in nested loops, it refers to the innermost loop containing that statement.

• The label \texttt{LoopCurrent} is visible in loop annotations and all other annotations related to a statement enclosed in a loop. It refers to the state at the beginning of the current step of the loop (see section 2.4.2 for more details on what constitutes a loop step in presence of side-effects in the condition). When \texttt{LoopCurrent} is used in a statement enclosed in nested loops, it refers to the innermost loop containing that statement.

Inside loop annotations, the labels \texttt{LoopCurrent} and \texttt{Here} are equivalent, except inside clauses \texttt{loop free} (see section 2.7.3) where \texttt{Here} is equivalent to \texttt{LoopEntry}.

There is one particular case for \texttt{assigns} and \texttt{ensures} clauses of function contracts where formal parameters of functions cannot refer to the label \texttt{Post}. In such clauses formal parameters always refer implicitly to the label \texttt{Pre}, and any \texttt{at} construct can modify the interpretation of formal parameters.

No logic label is visible in global logic declarations such as lemmas, axioms, definition of predicate or logic functions. When such an annotation needs to refer to a given memory state, it has to be given a label binder: this is described in Section 2.6.9.

\textbf{Example 2.28} The code below implements the famous extended Euclid’s algorithm for computing the greatest common divisor of two integers \texttt{x} and \texttt{y}, while computing at the same time two Bézout coefficients \texttt{p} and \texttt{q} such that $p \times x + q \times y = \gcd(x, y)$. The loop invariant for the Bézout property needs to refer to the value of \texttt{x} and \texttt{y} in the pre-state of the function.

```c
/*@ requires x >= 0 && y >= 0; */
@ behavior bezoutProperty:
@ ensures (*p)*x+(*q)*y == \result;
@*/
int extended_Euclid(int x, int y, int *p, int *q) {
    int a = 1, b = 0, c = 0, d = 1;
    /*@ loop invariant x >= 0 && y >= 0 ; */
    @ for bezoutProperty: loop invariant
    @  a*\at(x,Pre)+b*\at(y,Pre) == x &&
    @  c*\at(x,Pre)+d*\at(y,Pre) == y ;
```
```c
@ loop variant y;
@*/
while (y > 0) {
    int r = x % y;
    int q = x / y;
    int ta = a, tb = b;
    x = y; y = r;
    a = c; b = d;
    c = ta - c * q; d = tb - d * q;
}
*p = a; *q = b;
return x;
}
```

**Example 2.29** Here is a toy example illustrating tricky issues with `\at` and labels:

```c
int i;
int t[10];
//@ ensures 0 <= result <= 9;
int any();
//@ assigns i,t[\at(i,Post)];
//@ ensures t[i] == \old(t[\at(i,Here)]) + 1;
//@ ensures \let j = i; t[j] == \old(t[j]) + 1;
//@
void f() {
    i = any();
    t[i]++;
}
```

The two ensures clauses are equivalent. The simpler clause `t[i] == \old(t[i]) + 1` would be wrong because in `\old(t[i])`, `i` denotes the value of `i` in the pre-state.

Also, the assigns clause `i,t[i]` would be wrong too because again in `t[i]`, the value of `i` in the pre-state is considered.

**Example 2.30** Here is an example illustrating the use of LoopEntry and LoopCurrent

```c
void f (int n) {
    for (int i = 0; i < n; i++) {
        //@ assert \at(i,LoopInit) == 0; */
        int j = 0;
        while (j++ < i) {
            //@ assert \at(j,LoopInit) == 0; */
            //@ assert \at(j,LoopCurrent) + 1 == j; */
        }
    }
}
```

### 2.4.4 Statement contracts

The grammar for statement contracts is given in Figure 2.11. It is similar to function contracts, but without decreases clause. Additionally, a statement contract may refer to enclosing
named behaviors, with the form for id:... Such contracts are only valid for the corresponding behaviors, in particular only under the corresponding assumes clause.

The ensures clause does not constrain the post-state when the annotated statement is terminated by a goto jumping out of it, by a break, continue or return statement, or by a call to the exit function. To specify such behaviors, abrupt clauses (described in Section 2.9) need to be used.

On the other hand, it is different with assigns clauses. The locations having their value modified during the path execution, starting at the beginning of the annotated statement and leading to a goto jumping out of it, should be part of its assigns clause.

Example 2.31 The clause assigns \nothing; does not hold for that statement, even if the clause ensures x==\old(x); holds:

```
1 /*@ assigns x;  
2 @  ensures x==\old(x);  
3 @*/  
4 if (c) {  
5      x++;  
6     goto L;  
7   }  
8 L: ...  
```

Allocation-clauses allow to specify which memory location is dynamically allocated or deallocated by the annotated statement from the heap; they are defined later in Section 2.7.3.

2.5 Termination

The property of termination concerns both loops and recursive function calls. Termination is guaranteed by attaching a measure function to each loop (aspect already addressed in Section 2.4.2) and each recursive function. By default, a measure is an integer expression, and measures are compared using the usual ordering over integers (Section 2.5.1). It is also possible to define measures into other domains and/or using a different ordering relation (Section 2.5.2).
2.5. TERMINATION

2.5.1 Integer measures

Functions are annotated with integer measures with the syntax

```plaintext
//@ decreases e;
```

and loops are annotated similarly with the syntax

```plaintext
//@ loop variant e;
```

where the logic expression \( e \) has type \( \text{integer} \). For recursive calls, or for loops, this expression must decrease for the relation \( R \) defined by

```plaintext
R(x,y) <==> x > y && x >= 0.
```

In other words, the measure must be a decreasing sequence of integers which remain nonnegative, except possibly for the last value of the sequence (See example 2.24).

**Example 2.32** The clause `loop variant u-l;` can be added to the loop annotations of the example 2.23. The measure \( u-l \) decreases at each iteration, and remains nonnegative, except at the last iteration where it may become negative.

```plaintext
//@ ... 
//@ loop variant u-l; */ 
while ...
```

2.5.2 General measures

More general measures on other types can be provided, using the keyword `for`. For functions it becomes

```plaintext
//@ decreases e for R;
```

and for loops

```plaintext
//@ loop variant e for R;
```

In those cases, the logic expression \( e \) has some type \( \tau \) and \( R \) must be relation on \( \tau \), that is a binary predicate declared (see Section 2.6 for details) as

```plaintext
//@ predicate R(\tau x, \tau y) ... 
```

Of course, to guarantee termination, it must be proved that \( R \) is a well-founded relation.

**Example 2.33** The following example illustrates a variant annotation using a pair of integers, ordered lexicographically.

```plaintext
//@ ensures \result >= 0; 
int dummy();
//@ type intpair = (integer,integer);
/*@ predicate lexico(intpair p1, intpair p2) = 
  @ \let (x1,y1) = p1 ;
  @ \let (x2,y2) = p2 ;
  @ x1 < x2 && 0 <= x2 ||
  @ x1 == x2 && 0 <= y2 && y1 < y2;
```
2.5.3 Recursive function calls

The precise semantics of measures on recursive calls, especially in the general case of mutually recursive functions, is given as follows. We call cluster a set of mutually recursive functions which is a strongly connected component of the call graph. Within each cluster, each function must be annotated with a decreases clause with the same relation $R$ (syntactically). Then, in the body of any function $f$ of that cluster, any recursive call to a function $g$ must occur in a state where the measure attached to $g$ is smaller (w.r.t $R$) than the measure of $f$ in the pre-state of $f$. This also applies when $g$ is $f$ itself.

Example 2.34 Here are the classical factorial and Fibonacci functions:

```c
/*@ requires n <= 12;
@ decreases n;
@*/ int fact(int n) {
    if (n <= 1) return 1;
    return n * fact(n-1);
}

//@ decreases n;
int fib(int n) {
    if (n <= 1) return 1;
    return fib(n-1) + fib(n-2);
}
```

Example 2.35 This example illustrates mutual recursion:

```c
/*@ requires n>=0;
deCREASES n;
@*/ int even(int n) {
    if (n == 0) return 1;
    return odd(n-1);
}
```
2.6 Logic specifications

The language of logic expressions used in annotations can be extended by declarations of new logic types, and new constants, logic functions and predicates. These declarations follows the

2.5.4 Non-terminating functions

Experimental

There are cases where a function is not supposed to terminate. For instance, the main function of a reactive program might be a while(1) which indefinitely waits for an event to process. More generally, a function can be expected to terminate only if some preconditions are met. In those cases, a terminates clause can be added to the contract of the function, under the following form:

```c
//@ terminates p;
```

The semantics of such a clause is as follows: if p holds, then the function is guaranteed to terminate (more precisely, its termination must be proved). If such a clause is not present (and in particular if there is no function contract at all), it defaults to terminates \true; that is the function is supposed to always terminate, which is the expected behavior of most functions.

Note that nothing is specified for the case where p does not hold: the function may terminate or not. In particular, terminates \false; does not imply that the function loops forever. A possible specification for a function that never terminates is the following:

```c
/*@ ensures \false;
terminates \false;
*/
void f() { while(1); }
```

Example 2.36 A concrete example of a function that may not always terminate is the incr_list function of example 2.22. In fact, The following contract is also acceptable for this function:

```c
//@ this time, the specification accepts circular lists, but does not ensure that the function terminates on them (as a matter of fact, it does not).
//@ assigns { q->hd | struct list *q ; reachable(p,q) };
//@
void incr_list(struct list *p) {
    while (p) { p->hd++ ; p = p->next; }
}
```

2.6 Logic specifications

The language of logic expressions used in annotations can be extended by declarations of new logic types, and new constants, logic functions and predicates. These declarations follows the
2.6.1 Predicate and function definitions

New functions and predicates can be defined by explicit expressions, given after an equal sign.

Example 2.37 The following definitions

```plaintext
1  //@ predicate is_positive(integer x) = x > 0;
2  /*@ logic integer get_sign(real x) =
3     x > 0.0 ? 1 : ( x < 0.0 ? -1 : 0);
4  @*/
```
2.6 LOGIC SPECIFICATIONS

\[
\begin{align*}
\text{logic-def} &::= \text{inductive-def} \\
\text{inductive-def} &::= \text{inductive} \\
&\quad \text{poly-id parameters? \{ indcase* \}} \\
\text{indcase} &::= \text{case poly-id : pred ;}
\end{align*}
\]

Figure 2.13: Grammar for inductive definitions

illustrates the definition of a new predicate \texttt{is_positive} with an integer parameter, and a new logic function \texttt{sign} with a real parameter returning an integer.

2.6.2 Lemmas

Lemmas are user-given propositions, a facility that might help theorem provers to establish validity of ACSL specifications.

Example 2.38 The following lemma

\[
\begin{align*}
1 &\text{ /*@ lemma mean_property: } \forall \text{ integer } x,y; \ x \leq y \Longrightarrow x \leq (x+y)/2 \leq y; \\
2 &\text{ is a useful hint for program like binary search.}
\end{align*}
\]

Of course, a complete verification of an ACSL specification has to provide a proof for each lemma.

2.6.3 Inductive predicates

A predicate may also be defined by an inductive definition. The grammar for those style of definitions is given on Figure 2.13.

In general, an inductive definition of a predicate \( P \) has the form

\[
\begin{align*}
1 &\text{ /*@ inductive } P(x_1,\ldots,x_n) \{ \\
2 &\quad \text{ case } c_1 : p_1; \\
3 &\quad \ldots \\
4 &\quad \text{ case } c_k : p_k; \\
5 &\quad \} \text{ */}
\end{align*}
\]

where each \( c_i \) is an identifier and each \( p_i \) is a proposition.

The semantics of such a definition is that \( P \) is the least fixpoint of the cases, i.e. is the smallest predicate (in the sense that it is false the most often) satisfying the propositions \( p_1,\ldots,p_k \). With this general form, the existence of a least fixpoint is not guaranteed, so tools might enforce syntactic conditions on the form of inductive definitions. A standard syntactic restriction could be to allow only propositions \( p_i \) of the form

\[
\forall y_1,\ldots,y_m, h_1 \Longrightarrow \cdots \Longrightarrow h_l \Longrightarrow P(t_1,\ldots,t_n)
\]

where \( P \) occurs only positively in hypotheses \( h_1,\ldots,h_l \) (definite Horn clauses, \url{http://en.wikipedia.org/wiki/Horn_clause}).

Example 2.39 The following introduce a predicate \texttt{isgcd(x,y,d)} meaning that \( d \) is the greatest common divisor of \( x \) and \( y \).
CHAPTER 2. SPECIFICATION LANGUAGE

Figure 2.14: Grammar for axiomatic declarations

/*@ inductive is_gcd(integer a, integer b, integer d) {*/
@ case gcd_zero:
@  \forall integer n; is_gcd(n,0,n);
@ case gcd_succ:
@  \forall integer a,b,d; is_gcd(b, a % b, d) ==> is_gcd(a,b,d);
@ */

This definition uses definite Horn clauses, hence is consistent.

Example 2.22 already introduced an inductive definition of reachability in linked-lists, and was also bases on definite Horn clauses thus consistent.

2.6.4 Axiomatic definitions

Instead of an explicit definition, one may introduce an axiomatic definitions for a set of types, predicates and logic functions, which amounts to declare the expected profiles and a set of axioms. The grammar for those constructions is given on Figure 2.14.

Example 2.40 The following axiomatization introduce a theory of finite lists of integers a la LISP.

/*@ axiomatic IntList {*/
@  type int_list;
@  logic int_list nil;
@  logic int_list cons(integer n,int_list l);
@  logic int_list append(int_list l1,int_list l2);
Like inductive definitions, there is no syntactic conditions which would guarantee axiomatic definitions to be consistent. It is usually up to the user to ensure that the introduction of axioms does not lead to a logical inconsistency.

Example 2.41 The following axiomatization

```plaintext
/*@ a
logic integer get_sign(real x);
@ axiom sign_pos: \forall real x; x >= 0. ==> get_sign(x) == 1;
@ axiom sign_neg: \forall real x; x <= 0. ==> get_sign(x) == -1;
@ */
```

is inconsistent since it implies \(\text{sign}(0.0) == 1\) and \(\text{sign}(0.0) == -1\), hence \(-1 == 1\)

2.6.5 Polymorphic logic types

Experimental

We consider here an algebraic specification setting based on multi-sorted logic, where types can be polymorphic that is parametrized by other types. For example, one may declare the type of polymorphic lists as

```plaintext
//@ type list<A>;
```

One can then consider for instance list of integers (list <integer>), list of pointers (e.g. list <char*>), list of list of reals (list<list <real>>), etc.

The grammar of Figure 2.12 contains rules for declaring polymorphic types and using polymorphic type expressions.

2.6.6 Recursive logic definitions

Explicit definitions of logic functions and predicates can be recursive. Declarations in the same bunch of logic declarations are implicitly mutually recursive, so that mutually recursive functions are possible too.

Example 2.42 The following logic declaration

```plaintext
/*@ logic integer max_index{L}({int t[],integer n} =
@ (n==0) ? 0 :
@ (t[n-1]==0) ? n : max_index(t, n-1);
@ */
```

In this latter case, note that the two ‘>’ must be separated by a space, to avoid confusion with the shift operator.
defines a logic function which returns the maximal index $i$ between 0 and $n-1$ such that $t[i]=0$.

There is no syntactic condition on such recursive definitions, such as limitation to primitive recursion. In essence, a recursive definition of the form $f(\text{args}) = e$; where $f$ occurs in expression $e$ is just a shortcut for an axiomatic declaration of $f$ with an axiom $\forall \text{args}; f(\text{args}) = e$. In other words, recursive definitions are not guaranteed to be consistent, in the same way that axiomatics may introduce inconsistency. Of course, tools might provide a way to check consistency.

### 2.6.7 Higher-order logic constructions

**Experimental**

Figure 2.15 introduces new term constructs for higher-order logic.

**Abstraction** The term $\lambda \tau_1 x_1, \ldots, \tau_n x_n; t$ denotes the $n$-ary logic function which maps $x_1, \ldots, x_n$ to $t$. It has the same precedence as $\forall$ and $\exists$.

**Extended quantifiers** Terms $\text{quant}(t_1, t_2, t_3)$ where quant is max, min, sum, product or numof are extended quantifications. $t_1$ and $t_2$ must have type integer, and $t_3$ must be a unary function with an integer argument, and a numeric value (integer or real) except for numof for which it should have a boolean value. Their meanings are given as follows:

- $\text{max}(i, j, f) = \max\{f(i), f(i+1), \ldots, f(j)\}$
- $\text{min}(i, j, f) = \min\{f(i), f(i+1), \ldots, f(j)\}$
- $\text{sum}(i, j, f) = f(i) + f(i+1) + \cdots + f(j)$
- $\text{product}(i, j, f) = f(i) \times f(i+1) \times \cdots \times f(j)$
- $\text{numof}(i, j, f) = \#\{k | i \leq k \leq j \land f(k)\} = \sum(i, j, \lambda \text{integer } k ; f(k) ? 1 : 0)$

If $i>j$ then $\text{sum}$ and $\text{numof}$ above are 0, $\text{product}$ is 1, and $\text{max}$ and $\text{min}$ are unspecified (see Section 2.2.2).

**Example 2.43** Function that sums the element of an array of doubles.

```plaintext
/*@ requires n >= 0 && \valid(t+(0..n-1)) ; @ ensures \result == \sum(0,n-1,\lambda \text{integer } k ; t[k]) ; @*/
double array_sum(double t[],int n) {
int i;
double s = 0.0;
/* loop invariant 0 <= i <= n; @ loop invariant s == \sum(0,i-1,\lambda \text{integer } k ; t[k]) ; @ loop variant n-i; */
for(i=0; i < n; i++) s += t[i];
}```
2.6. LOGIC SPECIFICATIONS

logic-def ::= type logic-type =
  logic-type-def ;

logic-type-def ::= record-type | sum-type
  | type-expr

record-type ::= { type-expr id
  ( ; type-expr id)* ;?

sum-type ::= |? constructor
  ( | constructor)*

constructor ::= id
  | id
  ( type-expr
  (, type-expr)* )

  non-constant constructor

type-expr ::= ( type-expr
  (, type-expr)+ )

  product type

term ::= term . id
  | \match term
  { match-cases }
  | ( term (, term)+
  )
tuples

  | { (. id = term ;)*
  }
  records

  | \let ( id (, id)+ ) =
  term ; term

match-cases ::= match-case+

match-case ::= case pat : term

pat ::= id
  | id ( pat (, pat)*
  )
  non-constant constructor

  | pat | pat
  or pattern

  | -
  any pattern

  | literal | { (. id = pat)*
  }
  record pattern

  | ( pat (, pat)*
  )
  tuple pattern

  | pat as id
  pattern binding

Figure 2.16: Grammar for concrete logic types and pattern-matching

12 | return s;
13 | }

2.6.8 Concrete logic types

EXPERIMENTAL

Logic types may not only be declared but also be given a definition. Defined logic types can
be either record types, or sum types. These definitions may be recursive. For record types,
the field access notation t.id can be used, and for sum types, a pattern-matching construction
is available. Grammar rules for these additional constructions are given in Figure 2.16
CHAPTER 2  SPECIFICATION LANGUAGE

\[
\begin{align*}
\text{poly-id} & \ ::= \ id \quad \text{normal identifier} \\
                  & \ | \ id \ type-var-binders \quad \text{identifier for polymorphic object} \\
                  & \ | \ id \ label-binders \quad \text{normal identifier with labels} \\
                  & \ | \ id \ label-binders \quad \text{polymorphic identifier with labels} \\
\text{label-binders} & \ ::= \ \{ \ id \ (, \ id)^* \}
\end{align*}
\]

Figure 2.17: Grammar for logic declarations with labels

Example 2.44  The declaration

\[
//@ type list\langle A\rangle = Nil | Cons(A,list\langle A\rangle);
\]

introduces a concrete definition of finite lists. The logic definition

\[
//@ logic integer list_length\langle A\rangle(list\langle A\rangle l) = \\
@ \begin{align*}
@ & \text{match } l \{ \\
@ & \text{case } Nil : 0 \\
@ & \text{case } Cons(h,t) : 1+\text{list_length}(t)
\end{align*};
\]

defines the length of a list by recursion and pattern-matching.

2.6.9 Hybrid functions and predicates

Logic functions and predicates may take both (pure) C types and logic types arguments. Such an hybrid predicate (or function) can either be defined with the same syntax as before (or axiomatized).

Be it defined either directly by an expression or through a set of axioms, an hybrid function (or predicate) usually depends on one or more program points, because it depends upon memory states, via expressions such as:

- pointer dereferencing: *p, p->f;
- array access: t[i];
- address-of operator: &x;
- built-in predicate depending on memory: \text{\texttt{valid}}

To make such a definition safe, it is mandatory to add after the declared identifier a set of labels, between curly braces, as shown on Figure 2.17. Expressions as above must then be enclosed into the \texttt{at} construct to refer to a given label. However, to ease reading of such logic expressions, it is allowed to omit a label whenever there is only one label in the context.

Example 2.45  The following annotations declare a function which returns the number of occurrences of a given double in a memory block storing doubles between the given indexes, together with the related axioms. It should be noted that without labels, this axiomatization would be inconsistent, since the function would not depend on the values stored in \( t \), hence the two last axioms would say both that \( a==b+1 \) and \( a==b \) for some \( a \) and \( b \).
Example 2.46 This second example defines a predicate which indicates whether two memory blocks of the same size are a permutation of each other. It illustrates the use of more than a single label. Thus, the @ at operator is mandatory here. Indeed the two blocks may come from two distinct memory states. Typically, one of the post condition of a sorting function would be permut{Pre,Post}(t,t).

2.6.10 Memory footprint specification: reads clause

Experimental

Logic declaration can be augmented with a reads clause, with the syntax given in Figure 2.18, which extends the one of Figure 2.12. This feature allows to specify the footprint of an hybrid
### CHAPTER 2. SPECIFICATION LANGUAGE

**logic-function-decl** ::= logic type-expr poly-id parameters reads-clause ;

**logic-predicate-decl** ::= predicate poly-id parameters? reads-clause ;

**reads-clause** ::= reads locations

**logic-function-def** ::= logic type-expr poly-id parameters reads-clause = term ;

**logic-predicate-def** ::= predicate poly-id parameters? reads-clause = pred ;

---

**Figure 2.18: Grammar for logic declarations with reads clauses**

Predicate or function, that is the set of memory locations which it depends on. From such an information, one might deduce properties of the form \( f\{L_1\}(args) = f\{L_2\}(args) \) if it is known that between states \( L_1 \) and \( L_2 \), the memory changes are disjoint from the declared footprint.

**Example 2.47** The following is the same as example 2.45 augmented with a reads clause.

```plaintext
/*@ axiomatic Nb_occ {
  logic integer nb_occ{L}(double *t, integer i, integer j, double e) {
    reads t[i..j];
    axiom nb_occ_empty{L}: // ...
  }
  // ...
}@*/
```

If for example a piece of code between labels \( L_1 \) and \( L_2 \) modifies \( t[k] \) for some index \( k \) outside \( i..j \), then one can deduce that \( nb_occ{L_1}(t,i,j,e) = nb_occ{L_2}(t,i,j,e) \).

### 2.6.11 Specification Modules

Specification modules can be provided to encapsulate several logic definitions, for example

```plaintext
/*@ module List {
  type list<A> = Nil | Cons(A , list<A>);
  logic integer length<A>(list<A> l) = \match l { 
    case Nil : 0 
    case Cons(h,t) : 1+length(t) } ;
  logic A fold_right<A,B>((A -> B -> B) f, list<A> l, B acc) = \match l { 
    case Nil : acc 
    case Cons(h,t) : f(h,fold_right(f,t,acc)) } ;
}@*/
```
2.7. POINTERS AND PHYSICAL ADRESSING

Grammar for terms and predicates is extended with new constructs given in Figure 2.19.

Location-address denotes the address of a memory location. It is a set of terms of some pointer type as defined in Section 2.3.4.

2.7.1 Memory blocks and pointer dereferencing

C memory is structured into allocated blocks that can come either from a declarator or a call to one of the calloc, malloc or realloc functions. Blocks are characterized by their base address, i.e. the address of the declared object (the first declared object in case of an array declarator), or the pointer returned by the allocating function (when the allocation succeed) and their length.
ACSL provides the following built-in functions to deal with allocated blocks. All of them takes an optional label identifier as argument. The default value of that label is defined in Section 2.4.3.

- \( \text{base_addr}(L)(p) \) returns the base address of the allocated block containing, at the label L, the pointer p
  \[ \text{base_addr}\{id\} : \text{void}^* \rightarrow \text{char}^* \]
- \( \text{block_length}(L)(p) \) returns the length (in bytes) of the allocated block containing, at the label L, its argument pointer.
  \[ \text{block_length}\{id\} : \text{void}^* \rightarrow \text{size_t} \]

In addition, dereferencing a pointer may lead to run-time errors. A pointer p is said to be valid if \( *p \) is guaranteed to produce a definite value according to the C standard [12]. The following built-in predicates deal with this notion:

- \( \text{valid} \) applies to a set of terms (see Section 2.3.4) of some pointer type. \( \text{valid}(L)(s) \) holds if and only if dereferencing any \( p \in s \) is safe at label L, both for reading from \( *p \) and writing to it. In particular, \( \text{valid}(L)(\emptyset) \) holds for any label L.
  \[ \text{valid}\{id\} : \text{set}<\alpha^*> \rightarrow \text{boolean} \]
- \( \text{valid_read} \) applies to a set of terms of some pointer type and holds if and only if it is safe to read from all the pointers in the set
  \[ \text{valid_read}\{id\} : \text{set}<\alpha^*> \rightarrow \text{boolean} \]

\( \text{valid}(L)(s) \) implies \( \text{valid_read}(L)(s) \) but the reverse is not true. In particular, it is allowed to read from a string literal, but not to write in it (see [12], 6.4.5§6).

The status of \( \text{valid} \) and \( \text{valid_read} \) constructs depends on the type of their argument. Namely, \( \text{valid}(L)((\text{int}*)p) \) and \( \text{valid}(L)((\text{char}*)p) \) are not equivalent. On the other hand, if we ignore potential alignment constraints, the following equivalence is true for any pointer p:

\[
\text{valid}(L)(p) \iff \text{valid}(L)(((\text{char}*)p)+(0..\text{sizeof}(p)))
\]

and

\[
\text{valid_read}(L)(p) \iff \text{valid_read}(L)(((\text{char}*)p)+(0..\text{sizeof}(p)))
\]

Some shortcuts are provided:

- \( \text{null} \) is an extra notation for the null pointer (i.e. a shortcut for (void*)0). As in C itself (see [12], 6.3.2.3§3), the constant 0 can have any pointer type. In addition, \( \text{valid}(L)(\text{null}) \) is always false, for any logic label L. Of course, \( \text{valid_read}(L)(\text{null}) \) is always false too.

- \( \text{offset}(L)(p) \) returns the offset between p and its base address
  \[ \text{offset}\{id\} : \text{void}^* \rightarrow \text{size_t} \]
  \[ \text{offset}(L)(p) = ((\text{char}*)p - \text{base_addr}(L)(p)) \]

Again, if there is no alignment constraints, the following property holds: for any set of pointers \( s \) and label L, \( \text{valid_read}(L)(s) \) if and only if for all \( p \in s \):

\[
\text{offset}(L)(p) \geq 0 \land \text{offset}(L)(p) + \text{sizeof}(p) \leq \text{block_length}(L)(p)
\]
2.7. POINTERS AND PHYSICAL ADRESSING

| allocation-clause ::= | allocates dyn-allocation-addresses ; |
|                     | frees dyn-allocation-addresses ; |
| loop-allocation ::= | loop allocates dyn-allocation-addresses ; |
|                     | loop frees dyn-allocation-addresses ; |
| dyn-allocation-addresses ::= | location-addresses |
|                            | | nothing |

Figure 2.20: Grammar for dynamic allocations and deallocations

2.7.2 Separation

ACSL provides a built-in function to deal with separation of locations:

- \texttt{\separated} applies to sets of terms (see Section 2.3.4) of some pointer type. \texttt{\separated(s_1,s_2)} holds for any set of pointers \(s_1\) and \(s_2\) if and only if for all \(p \in s_1\) and \(q \in s_2\):

\[
\forall i,j; 0 \leq i < \text{sizeof}(*p), 0 \leq j < \text{sizeof}(*q) \implies \text{(char*)} p + i \neq \text{(char*)} q + j
\]

In fact, \texttt{\separated} is an \(n\)-ary predicate. \texttt{\separated(s_1,\ldots,s_n)} means that for each \(i \neq j\), \texttt{\separated(s_i,s_j)}.

2.7.3 Dynamic allocation and deallocation

Experimental


Allocation-clauses allow to specify which memory location is dynamically allocated or deallocated. The grammar for those constructions is given on Figure 2.20.

\texttt{allocates} \texttt{\nothing} and \texttt{frees} \texttt{\nothing} are respectively equivalent to \texttt{allocates} \texttt{\empty} and \texttt{frees} \texttt{\empty}; it is left for convenience like for \texttt{assigns} clauses.

Allocation clauses for function and statement contracts

Clauses \texttt{allocates} and \texttt{frees} are tied together. The simple contract

\[
/*@ frees P_1,P_2,\ldots; @*/
\]

\[
/*@ allocates Q_1,Q_2,\ldots; @*/
\]

means that any memory address that does not belong to the union of sets of terms \(P_i\) and \(Q_j\) has the same \textit{allocation status} (see below) in the post-state than in the pre-state. The only one difference between \texttt{allocates} and \texttt{frees} is that sets \(P_i\) are evaluated in the pre-state, and sets \(Q_i\) are evaluated in the post-state.

The built-in type \texttt{allocation_status} can take the following values:

\[
/*@
\texttt{type allocation\_status =}
\texttt{\static | \register | \automatic | \dynamic | \unallocated;}@*/
\]
Built-in function \( \text{allocation}\{L\}(p) \) returns the allocation status of the block containing, at the label \( L \), the pointer \( p \).

\[ \text{allocation}\{id\} : \text{void*} \rightarrow \text{allocation\_status} \]

This function is such that for any pointer \( p \) and label \( L \)

\[ \text{allocation}\{L\}(p) = \text{allocation}\{\text{base\_addr}(p)\} \]

and

\[ \text{allocation}\{L\}(p) = \text{unallocated} \implies \neg \text{valid\_read}\{L\}(p) \]

allocates \( Q_1, ..., Q_n \) is equivalent to the postcondition

\[
\text{forall char* p; } \\
\text{separated } (\text{union}(Q_1, ..., Q_n), p) ==> \\
(\text{base\_addr}\{\text{Here}(p)\} == \text{base\_addr}\{\text{Pre}(p)\}) \\
\land \\
(\text{block\_length}\{\text{Here}(p)\} == \text{block\_length}\{\text{Pre}(p)\}) \\
\land \\
(\text{valid}\{\text{Here}(p)\} <=\text{valid}\{\text{Pre}(p)\}) \\
\land \\
(\text{valid\_read}\{\text{Here}(p)\} <=\text{valid\_read}\{\text{Pre}(p)\}) \\
\land \\
(\text{allocation}\{\text{Here}(p)\} == \text{allocation}\{\text{Pre}(p)\})
\]

In fact, like the assigns clause does not specify which memory location is assigned, the allocation-clauses do not specify which memory location is dynamically allocated or deallocated. Pre-conditions and post-conditions should be added to complete specifications about allocations and deallocations. The following shortcuts can be used for that:

- \( \text{allocable}\{L\}(p) \) holds if and only if the pointer \( p \) refers, at the label \( L \), to the base address of an unallocated memory block.
  \[ \text{allocable}\{id\} : \text{void*} \rightarrow \text{boolean} \]
  For any pointer \( p \) and label \( L \)

\[ \text{allocable}\{L\}(p) \iff (\text{allocation}\{L\}(p) == \text{unallocated} \land p == \text{base\_addr}\{L\}(p)) \]

- \( \text{freeable}\{L\}(p) \) holds if and only if the pointer \( p \) refers, at the label \( L \), to an allocated memory block that can be safely released using the C function \texttt{free}.
  \[ \text{freeable}\{id\} : \text{void*} \rightarrow \text{boolean} \]
  For any pointer \( p \) and label \( L \)

\[ \text{freeable}\{L\}(p) \iff (\text{allocation}\{L\}(p) == \text{dynamic} \land p == \text{base\_addr}\{L\}(p)) \]

- \( \text{fresh}\{L_0, L_1\}(p, n) \), indicates that \( p \) refers to an allocated memory block at label \( L_1 \), but that it is not the case at label \( L_0 \). The predicate ensures also that, at label \( L_1 \), the length (in bytes) of the block allocated dynamically equals to \( n \).
  \[ \text{fresh}\{id, id\} : \text{void*}, \text{integer} \rightarrow \text{boolean} \]
  For any pointer \( p \) and labels \( L_0 \) and \( L_1 \)

\[
\text{fresh}\{L_0, L_1\}(p, n) \iff (\text{allocation}\{L_0\}(p) \land \text{freeable}\{L_1\}(p) \land \text{block\_length}\{L_1\}(p) == n \land \text{valid}\{L_1\}((\text{char*})p + (0 .. (n-1))))
\]

Example 2.48 malloc and free functions can be specified as follows.
typedef unsigned long size_t;

/*@ assigns \nothing;
@ allocates \result;
@ ensures \result == \null || \fresh\{Old,Here\}(\result,n);
@*/
void *malloc(size_t n);

/*@ requires p!=\null ==> \freeable\{Here\}(p);
@ assigns \nothing;
@ frees p;
@ ensures p!=\null ==> \allocable\{Here\}(p);
@*/
void free(void *p);

Default labels for constructs dedicated to memory are such that logic label Here can be omitted.

When a behavior contains only one of the two allocation clauses, the given clause specifies the whole set of memory addresses to consider. This means that the set value for the other clause of that behavior defaults to \nothing. Now, when none of the two allocation clauses is given, the meaning is different for anonymous behaviors and named behaviors:

- a named behavior without allocation clause does not specify anything about allocations and deallocations. The allocated and deallocated memory blocks are in fact specified by the anonymous behavior of the contract. There is no condition to verify for these named behaviors about allocations and deallocations;
- for an anonymous behavior, no allocation clause means that there is no newly allocated nor deallocated memory block. That is equivalent to give allocates \nothing.

These rules are such that contracts without any allocation clause should be considered as having only one allocates \nothing; leading to a condition to verify for each anonymous behavior.

**Example 2.49** More precise specifications can be given using named behaviors under the assumption of assumes clauses.
The behaviors named allocation and deallocation do not need an allocation clause. For example, the allocation constraint of the allocation behavior is given by the clause allocates \result \nothing of the anonymous behavior of the malloc function contract. To set a stronger constraint into the behavior named no_allocation, the clause allocates \nothing should be given.

Allocation clauses for loop annotations

Loop annotations are complemented by similar clauses allowing to specify which memory location is dynamically allocated or deallocated by a loop. The grammar for those constructions is given on Figure 2.20.

The clauses loop allocates and loop frees are tied together. The simple loop annotation

```plaintext
/*@ loop allocates Q1, Q2, ...; */
```

means that any memory address that does not belong to the union of sets of terms $P_i$ and $Q_i$ has the same allocation status in the current state than before entering the loop. The only difference between these two clauses is that sets $P_i$ are evaluated in the state before entering the loop (label LoopEntry), and $Q_i$ are evaluated in the current loop state (label LoopCurrent).

Namely, as for loop assigns, loop annotations loop frees and loop allocates define a loop invariant.

More precisely, the following loop annotation:

```plaintext
//@ loop allocates Q1, Q2, ...; */
```

is equivalent to the loop invariant:

```plaintext
\forall char* p; 
\text{separated}(\text{union}(Q1, \ldots, Q_n), p) =>
(\text{base_addr}(\text{LoopEntry})(p)==\text{base_addr}(\text{Here})(p) && \text{block_length}(\text{Here})(p)==\text{block_length}(\text{LoopEntry})(p))
```
2.8. SETS AS FIRST-CLASS VALUES

Example 2.50
/*@ assert \forall integer j; 0<=j<n ==> \freeable(q[j]); */
/*@ loop assigns q[0..(i-1)]; */
@ loop freeing q[0..\at(i-1,LoopCurrent)];
@ loop invariant \forall integer j ; 0 <= j < i ==> \allocable(\at(q[j],LoopEntry));
@ loop invariant \forall integer j ; 0 <= i <= n;
@*/
for (i=0; i<n; i++) {
  free(q[i]);
  q[i]=NULL;
}

The addresses of locations \texttt{q[0..n]} are not modified by the loop, but their values are. The clause \texttt{loop frees} catches the set of the memory blocks that may have been released by the previous loop iterations. The first \texttt{loop invariant} defines exactly these memory blocks. On the other hand, \texttt{loop frees} indicates that the remaining blocks have not been freed since the beginning of the loop. Hence, they are still \texttt{\freeable} as expressed by the initial \texttt{assert}, and \texttt{free(q[i])} will succeed at next step.

A \texttt{loop-clause} without allocation clause implicitly contains \texttt{loop allocates nothing}. That means the \texttt{allocation status} is not modified by the loop body. A \texttt{loop-behavior} without allocation clause means that the allocated and deallocated memory blocks are in fact specified by the allocation clauses of the \texttt{loop-clauses} (Grammar of \texttt{loop-clauses} and \texttt{loop-behaviors} is given in Figure 2.9).

2.8 Sets as first-class values

Sets of terms, as defined in Section 2.3.4, can be used as first-class values in annotations. All the elements of such a set must share the same type (modulo the usual implicit conversions). Sets have the built-in type \texttt{set<\texttt{A}>} where \texttt{A} is the type of terms contained in the set.

In addition, it is possible to consider sets of pointers to values of different types. In this case, the set is of type \texttt{set<char*>>} and each of its elements \texttt{e} is converted to \texttt{(char*)e + (0..sizeof(*e)-1)}.

Example 2.51 The following example defines the footprint of a structure, that is the set of locations that can be accessed from an object of this type.

```c
struct S {
  char *x;
  int *y;
};
//@ logic set<char*> footprint(struct S s) = \union(s.x,s.y);
```
**Figure 2.21: Grammar of contracts about abrupt terminations**

```
abrupt-clause-fn ::= exits-clause
exits-clause ::= exits pred ;
abrupt-clause-stmt ::= exits-clause |
                     breaks-clause |
                     continues-clause |
                     returns-clause
breaks-clause ::= breaks pred ;
continues-clause ::= continues pred ;
returns-clause ::= returns pred ;
term ::= \exit_status
```

In the first definition, since union is made with a `set<char*>` and a `set<int*>`, the result is a `set<char*>` (accordingly to typing of union). In other words, the two definitions above are equivalent.

This logic function can be used as argument of `\separated` or `assigns` clause.

Thus, the `\separated` predicate satisfies the following property (with `s_1` of type `set<\tau_1*>` and `s_2` of type `set<\tau_2*>`)

```
\separated(s_1, s_2) <==>
( \forall \tau_1* p; \forall \tau_2* q;
  \subset(p, s_1) \&\& \subset(q, s_2) ==> 
  ( \forall \integer i, j;
    0 <= i < \sizeof(\tau_1) \&\& 0 <= j < \sizeof(\tau_2) ==> 
    (char*)p + i != (char*)q + j))
```

and a clause `assigns L_1,...,L_n` is equivalent to the postcondition

```
( \forall char* p; \separated(\union(\&L_1,...,\&L_n), p) ==> *p == \old(*p)
```

### 2.9 Abrupt termination

**Experimental**

The `ensures` clause of function and statement contracts does not constrain the post-state when the annotated function and statement terminates respectively abruptly. In such cases, `abrupt clauses` can be used inside `simple clause` or `behavior body`. The allowed constructs are shown in Figure 2.21.
2.9. ABRUPT TERMINATION

The clauses breaks, continues and returns can only be found in a statement contract and state properties on the program state which hold when the annotated statement terminates abruptly with the corresponding statement (break, continue or return).

Inside these clauses, the construct \( \text{old}(e) \) is allowed and denotes, like for statement contracts ensures, assigns and allocates, the value of \( e \) in the pre-state of the statement. More generally, the visibility in abrupt clauses of predefined logics labels (presented in Section 2.4.3) is the same as in ensures clauses.

For the returns case, the \( \text{result} \) construct is allowed (if the function is not returning void) and is bound to the returned value.

**Example 2.52** The following example illustrates each abrupt clause of a statement contracts.

```plaintext
int f(int x) {
    while (x > 0) {

        /*\@
        @ breaks x % 11 == 0 && x == \text{old}(x);
        @ continues (x+1) % 11 != 0 && x % 7 == 0 && x == \text{old}(x)-1;
        @ returns (\text{result}+2) % 11 != 0 && x % 7 == 0
        @ \text{old}(x)-2;
        @ ensures (x+3) % 11 != 0 && x % 7 != 0 && (x+1) % 5 != 0
        @ x == \text{old}(x)-3;
        */
    {
        if (x % 11 == 0) break;
        x--;
        if (x % 7 == 0) continue;
        x--;
        if (x % 5 == 0) return x;
        x--;
    }
    return x;
}
```

The exits clause can be used in both function and statement contracts to give behavioral properties to the main function or to any function that may exit the program, e.g. by calling the exit function.

In such clauses, \( \text{old}(e) \) is allowed and denotes the value of \( e \) in the pre-state of the function or statement, and \( \text{exit_status} \) is bound to the return code, e.g. the value returned by main or the argument passed to exit. The construct \( \text{exit_status} \) can be used only in exits, assigns and allocates clauses. On the contrary, \( \text{result} \) cannot be used in exits clauses.

**Example 2.53** Here is a complete specification of the exit function which performs an unconditional exit of the main function:

```plaintext
/*@ assigns \nothing;
@ ensures \false;
@ exits \text{exit_status} == status;
*/
void exit(int status);
```
**CHAPTER 2. SPECIFICATION LANGUAGE**

| assigns-clause ::= assigns locations (\from locations) ; |
|                 | assigns term \from locations = term ; |

Figure 2.22: Grammar for dependencies information

```c
int status;
/*@ assigns status;
@ exits !cond && \exit_status == 1 && status == val;
*/
void may_exit(int cond, int val) {
    if (! cond) {
        status = val;
        exit(1);
    }
}
```

Note that the specification of the `may_exit` function is incomplete since it allows modifications of the variable `status` when no exit is performed. Using behaviors, it is possible to distinguish between the exit case and the normal case, as in the following specification:

```c
/*@ behavior no_exit :
@ assumes cond;
@ assigns \nothing;
@ exits \false;
@ behavior no_return :
@ assumes !cond;
@ assigns status;
@ exits \exit_status == 1 && status == val;
@ ensures \false;
*/
void may_exit(int cond, int val) ;
```

Contrary to `ensures` clauses, `assigns`, `allocates` and `frees` clauses of function and statement contracts constrain the post-state even when the annotated function and statement terminates respectively abruptly. This is shown in example 2.53 for a function contract.

### 2.10 Dependencies information

**Experimental**

An extended syntax of `assigns` clauses, described in Figure 2.22 allows to specify data dependencies and *functional expressions*.

Such a clause indicates that the assigned values can only depend upon the locations mentioned in the `\from` part of the clause. Again, this is an over-approximation: all of the locations involved in the computation of the modified values must be present, but some of locations might not be used in practice. If the `\from` clause is absent, all of the locations reachable at the given point of the program are supposed to be used. Moreover, for a single location, it is possible to give the precise relation between its final value and the value of its dependencies. This expression is evaluated in the pre-state of the corresponding contract.
Example 2.54  The following example is a variation over the array_sum function in example 2.43, in which the values of the array are added to a global variable total.

```c
double total = 0.0;
/*@ requires n >= 0 && \valid(t+(0..n-1)) ; @*/
void array_sum(double t[],int n) {
  int i;
  for(i=0; i < n; i++) total += t[i];
  return;
}
```

Example 2.55  The composite element modifier operators can be useful for writing such functional expressions.

```c
struct buffer { int pos ; char buf[80] ; } line;
/*@ requires 80 > line.pos >= 0 ; @*/
void add_eol() {
  line.buf[line.pos] = '\0';
}
```

2.11  Data invariants

Data invariants are properties on data that are supposed to hold permanently during the lifetime of these data. In ACSL, we distinguish between:

- **global** invariants and **type** invariants: the former only apply to specified global variables, whereas the latter are associated to a static type, and apply to any variables of the corresponding type;

- **strong** invariants and **weak** invariants: strong invariants must be valid at any time during program execution (more precisely at any sequence point as defined in the C standard), whereas weak invariants must be valid at function boundaries (function entrance and exit) but can be violated in between.

The syntax for declaring data invariants is given in Figure 2.23. The strength modifier defaults to weak.

Example 2.56  In the following example, we declare

1. a weak global invariant a_is_positive which specifies that global variable a should remain positive (weakly, so this property might be violated temporarily between function calls);
2. **a strong type invariant for variables of type** `temperature`;

3. **a weak type invariant for variables of type** `struct S`.

```
int a;
//@ global invariant a_is_positive: a >= 0 ;

typedef double temperature;
//@ strong type invariant temp_in_celsius(temperature t) =
@ t >= -273.15 ;
@*

struct S {
  int f;
};
//@ type invariant S_f_is_positive(struct S s) = s.f >= 0 ;
```

### 2.11.1 Semantics

The distinction between strong and weak invariants has to do with the sequence points where the property is supposed to hold. The distinction between global and type invariants has to do with the set of values on which they are supposed to hold.

- **Weak global invariants** are properties which apply to global data and hold at any function entrance and function exit.

- **Strong global invariants** are properties which apply to global data and hold at any step during execution (starting after initialization of these data).

- A **weak type invariant on type** `τ` **must hold at any function entrance and exit, and applies to any global variable or formal parameter with static type** `τ`. If the result of the function is of type `τ`, the result must also satisfy its weak invariant at function exit. However, it says nothing about fields, array elements, memory locations, etc. of type `τ`.

- A **strong type invariant on type** `τ` **must hold at any step during execution, and applies to any global variable, local variable, or formal parameter with static type** `τ`. If the result of the function has type `τ`, the result must also satisfy its strong invariant at function exit. Again, it says nothing about fields, array elements, memory locations, etc. of type `τ`.
Example 2.57 The following example illustrates the use of a data invariant on a local static variable.

```c
void out_char(char c) {
    static int col = 0;
    //@ global invariant I : 0 <= col <= 79;
    col++;
    if (col >= 80) col = 0;
}
```

Example 2.58 Here is a longer example, the famous Dijkstra’s Dutch flag algorithm.

```c
typedef enum { BLUE, WHITE, RED } color;
/*@ type invariant isColor(color c) =
@  c == BLUE || c == WHITE || c == RED ;
@*/

/*@ predicate permut(L1,L2)(color *t1, color *t2, integer n) =
@  @valid(t1+(0..n)),L1) && @valid(t2+(0..n)),L2) &&
@  @numof(0,n,\lambda integer i; \at(t1[i],L1) == BLUE) ==
@  @numof(0,n,\lambda integer i; \at(t2[i],L2) == BLUE)
@  @
@  @numof(0,n,\lambda integer i; \at(t1[i],L1) == WHITE) ==
@  @numof(0,n,\lambda integer i; \at(t2[i],L2) == WHITE)
@  @
@  @numof(0,n,\lambda integer i; \at(t1[i],L1) == RED) ==
@  @numof(0,n,\lambda integer i; \at(t2[i],L2) == RED);
@*/

/*@ requires @valid(t+i) && @valid(t+j);
@ assigns t[i],t[j];
@ ensures t[i] == \old(t[j]) && t[j] == \old(t[i]);
@*/
void swap(color t[], int i, int j) {
    int tmp = t[i];
    t[i] = t[j];
    t[j] = tmp;
}

typedef struct flag {
    int n;
    color *colors;
} flag;
/*@ type invariant is_colored(flag f) =
@  f.n >= 0 && @valid(f.colors+(0..f.n-1)) &&
@  @forall integer k; 0 <= k < f.n ==> isColor(f.colors[k]) ;
@*/

/*@ predicate isMonochrome(L)(color *t, integer i, integer j,
@  color c) =
@  @forall integer k; i <= k <= j ==> t[k] == c ;
@*/

/*@ assigns f.colors[0..f.n-1];
@ ensures
@  @\exists integer b, integer r;
```
2.11.2 Model variables and model fields

A model variable is a variable introduced in the specification with the keyword model. Its type must be a logic type. Analogously, types may have model fields. These are used to provide abstract specifications to functions whose concrete implementation must remain private.

The precise syntax for declaring model variables and fields is given in Figure 2.24. It is presented as additions to the regular C grammar for declarations.

Informal semantics of model variables is as follows.

- Model variables can only appear in specifications. They are not lvalues, thus they

| declaration ::= C-declaration |
| | /*@ model parameter ; */ model variable |
| | /*@ model C-type-name { parameter ; } ; */ model field |

Figure 2.24: Grammar for declarations of model variables and fields
cannot be assigned directly (unlike ghost variables, see below).

- Nevertheless, a function contract might state that a model variable is assigned.
- When a function contract mentions model variables:
  - the precondition is implicitly existentially quantified over those variables;
  - the postconditions are universally quantified over the old values of model variables,
    and existentially quantified over the new values.

Thus, in practice, the only way to prove that a function body satisfies a contract with model variables is to provide an invariant relating model variables and concrete variables, as in the example below.

Model fields behave the same, but they are attached to any value whose static type is the one of the model declaration. A model field can be attached to any C type, not only to `struct`. When it is attached to a compound type, however, it must not have the same name as a C field of the corresponding type. In addition, model fields are “inherited” by a `typedef` in the sense that the newly defined type has also the model fields of its parents (and can acquire more, which will not be present for the parent). For instance, in the following code, `t1` has one model field `m1`, while `t2` has two model fields, `m1` and `m2`.

```
1 typedef int t1;
2 typedef t1 t2;
3 /*@ model t1 { int m1 }; */
4 /*@ model t2 { int m2 }; */
```

**Example 2.59** Here is an example of a specification for a function which generates fresh integers. The contract is given in term of a model variable which is intended to represent the set of “forbidden” values, e.g. the values that have already been generated.

```
/* public interface */

1 //@ model set<integer> forbidden = \empty;
2
3 /*@ assigns forbidden; */
4 @ ensures \! \subset (\result,\old(forbidden))
5 @ \&\& \subset (\result,forbidden) \&\& \subset (\old(forbidden),forbidden);
6 */
7 int gen();
```

The contract is expressed abstractly, telling that

- the forbidden set of values is modified;
- the value returned is not in the set of forbidden values, thus it is “fresh”;
- the new set of forbidden values contains both the value returned and the previous forbidden values.

An implementation of this function might be as follows, where a decision has been made to generate values in increasing order, so that it is sufficient to record the last value generated. This decision is made explicit by an invariant.
/* implementation */

```c
int gen() {
    static int x = 0;
    /*@ global invariant I: \forall integer k; Set::mem(k, forbidden) ==> x > k; */
    return x++;
}
```

**Remarks** Although the syntax of model variables is close to JML model variables, they differ in the sense that the type of a model variable is a logic type, not a C type. Also, the semantics above is closer to the one of B machines [1]. It has to be noticed that program verification with model variables does not have a well-established theoretical background [19, 17], so we deliberately do not provide a precise semantics in this document.

### 2.12 Ghost variables and statements

Ghost variables and statements are like C variables and statements, but visible only in the specifications. They are introduced by the `ghost` keyword at the beginning of the annotation (i.e. `/*@ ghost ... */` or `//@ ghost ...` for a one-line ghost code, as mentioned in section 1.2). The grammar is given in Figure 2.25, in which only the first form of annotation is used. In this figure, the `C-*` non-terminals refer to the corresponding grammar rules of the ISO standard, without any ACSL extension. Any non terminal of the form `ghost-non-term` for which no definition is given in the figure represents the corresponding `C-non-term` entry, in which any `entry` is substituted by `ghost-entry`.

The variations with respect to the C grammar are the following:

- Comments must be introduced by `//` and extend until the end of the line (the ghost code itself is placed inside a C comment. `/* ... */` would thus lead to incorrect C code).

- It is however possible to write multi-line annotations for ghost code. These annotations are enclosed between `/*` and `*/`. As in normal annotations, `@`s at the beginning of a line and at the end of the comment (before the final `*/`) are considered as blank.

- Logical types, such as `integer` or `real` are authorized in ghost code.

- A non-ghost function can take ghost parameters. If such a ghost clause is present in the declarator, then the list of ghost parameters must be non-empty and fixed (no vararg `ghost`). The call to the function must then provide the appropriate number of ghost parameters.

- Any non-ghost `if-statement` which does not have a non-ghost `else` clause can be augmented with a ghost one. Similarly, a non-ghost `switch` can have a `default :` clause if it does not have a non-ghost one (there are however semantical restrictions for valid ghost labelled statements in a switch, see next paragraph for details).
2.12. GHOST VARIABLES AND STATEMENTS

| ghost-type-specifier ::= C-type-specifier |
| logic-type |
| declaration ::= C-declaration |
| /*@ ghost declaration */ |
| direct-declarator ::= C-direct-declarator |
| direct-declarator ( C-parameter-type-list? ) |
| /*@ ghost ( ghost-parameter-type-list ) */ |
| postfix-expression ::= C-postfix-expression |
| postfix-expression ( C-argument-expression-list? ) |
| /*@ ghost ( ghost-argument-expression-list ) */ |
| statement ::= C-statement |
| statements-ghost |
| statements-ghost ::= /*@ ghost statements-ghost */ |
| ghost-selection-statement ::= C-selection-statement |
| if ( C-expression ) |
| /*@ ghost else ghost-statement+ */ |
| struct-declaration ::= C-struct-declaration |
| /*@ ghost struct-declaration */ |

Figure 2.25: Grammar for ghost statements

Semantics of Ghost Code  The question of semantics is essential for ghost code. Informally, the semantics requires that ghost statements do not change the regular program execution. This implies several conditions, including e.g:

- Ghost code cannot modify a non-ghost C variable.
- Ghost code cannot modify a non-ghost structure field.
- If \( p \) is a ghost pointer pointing to a non-ghost memory location, then it is forbidden to assign \( *p \).
- Body of a ghost function is ghost code, hence do not modify non-ghost variables or fields.
• If a non-ghost C function is called in ghost code, it must not modify non-ghost variables or fields.

• If a structure has ghost fields, the sizeof of the structure is the same as the structure without ghost fields. Also, alignment of fields remains unchanged.

• The control-flow graph of a function must not be altered by ghost statements. In particular, no ghost return can appear in the body of a non-ghost function. Similarly, ghost goto, break, and continue cannot jump outside of the innermost non-ghost enclosing block.

Semantics is specified as follows. First, the execution of a program with ghost code involves a ghost memory heap and a ghost stack, disjoint from the regular heap and stack. Ghost variables lie in the ghost heap, so as the ghost field of structures. Thus, every memory side-effect can be classified as ghost or non-ghost. Then, the semantics is that memory side-effects of ghost code must always be in the ghost heap or the ghost stack.

Notice that this semantics is not statically decidable. It is left to tools to provide approximations, correct in the sense that any code statically detected as ghost must be semantically ghost.

Example 2.60 The following example shows some invalid assignments of ghost pointers:

```c
void f(int x, int *q) {
    //@ ghost int *p = q;
    //@ ghost *p = 0;
    // above assignment is wrong: it modifies *q which lies
    // in regular memory heap

    //@ ghost p = &x;
    //@ ghost *p = 0;
    // above assignment is wrong: it modifies x which lies
    // in regular memory stack
}
```

Example 2.61 The following example shows some invalid ghost statements:

```c
int f(int x, int y) {
    //@ ghost int z = x + y;
    switch (x) {
        case 0: return y;
        //@ ghost case 1: z=y;
        // above statement is correct.
        //@ ghost case 2: { z++; break; }
        // invalid, would bypass the non-ghost default
        default: y++;
    }
    return y;
}

int g(int x) {
    //@ ghost int z = x;
    if (x > 0) { return x; }
}
```
2.12. GHOST VARIABLES AND STATEMENTS

Differences between model variables and ghost variables A ghost variable is an additional specification variable which is assigned in ghost code like any C variable. On the other hand, a model variable cannot be assigned, but one can state it is modified and can express properties about the new value, in a non-deterministic way, using logic assertions and invariants. In other words, specifications using ghost variable assignments are executable.

Example 2.62 The example 2.59 can also be specified with a ghost variable instead of a model variable:

```c
//@ ghost set<
integer>
forbidden = \empty;
//@ assigns forbidden;
//@ ensures ! \subset(\result,\old(forbidden))
//@ \&\& \subset(\result,forbidden)
//@ \&\& \subset(\old(forbidden),forbidden);
//@
int
gen() {
  static int x = 0;
  //@ global invariant I: \forall integer k;
  @
  \subset(k,forbidden) ==> x > k;
  @*/
x++;
//@ ghost forbidden = \union(x,forbidden);
return x;
}
```

2.12.1 Volatile variables

Experimental

Volatile variables can not be used in logic terms, since reading such a variable may have a side effect, in particular two successive reads may return different values.

```
declaration ::= //@ volatile locations (reads id)? (writes id)?
```

Figure 2.26: Grammar for volatile constructs

Specifying properties of a volatile variable may be done via a specific construct to attach two ghost functions to it. This construct, described by the grammar of Figure 2.26, has the following shape:

```c
volatile \tau x;
//@ volatile x reads f writes g;
```

where f and g are ghost functions with the following prototypes:

```c
\tau f(\volatile \tau* p);
\tau g(\volatile \tau* p, \tau v);
```
This must be understood as a special construct to instrument the C code, where each access to the variable \( x \) is replaced by a call to \( f(&x) \), and each assignment to \( x \) of a value \( v \) is replaced by \( g(&x, v) \). If a given volatile variable is only read or only written to, the unused accessor function can be omitted from the volatile construct.

**Example 2.63** The following code is instrumented in order to inject fixed values at each read of variable \( x \), and collect written values.

```
volatile int x;

//@ ghost //@ requires p == &x;
int reads_x(volatile int *p) {
  static int injector_x[] = { 1, 2, 3 };
  static int injector_count = 0;
  if (p == &x)
    return injector_x[injector_count++];
  else
    return 0; // should not happen
}

//@ ghost
int reads
reads_x
writes
writes_x;

//@ ghost int collector_x[3];
//@ ghost int collector_count = 0;
//@ ghost //@ requires p == &x;
int writes_x(volatile int *p, int v) {
  if (p == &x)
    return collector_x[collector_count++] = v;
  else
    return 0; // should not happen
}

//@
volatile x reads
reads_x writes
writes_x;
//@
//@ ensures collector_count == 3 && collector_x[2] == 2;
//@ ensures \result == 6;
@/
int main () {
  int i, sum = 0;
  for (i=0 ; i < 3; i++) {
    sum += x;
    x = i;
  }
  return sum;
}
```

### 2.13 Undefined values, dangling pointers

#### 2.13.1 Initialization

\( \text{initialized} \{L\}(p) \) is a predicate taking a set of some pointer to l-values as argument and means that each l-value in this set is initialized at label \( L \).

\( \text{initialized} \{id\} : \text{set} <\alpha*> \rightarrow \text{bool} \)
Example 2.64 *In the following, the assertion is true.*

```c
def f(int n) {
    int x;
    if (n > 0) x = n; else x = -n;
    // @ assert \texttt{\textbackslash initialized \{Here\}(\&x)};
    return x;
}
```

Default labels are such that logic label \texttt{\{Here\}} can be omitted.

### 2.13.2 Dangling pointers

\texttt{\dangling\{L\}} is a predicate taking a set of some pointer to l-values as argument and means that each l-value in this set has a *dangling content* at label \texttt{L}. That, its value is (or contains bits of) a dangling address: either the address of a local variable refers to outside of its scope, or the address of a variable that has been dynamically allocated, then deallocated.

\texttt{\dangling\{id\} : set<\alpha*> \rightarrow bool}

Example 2.65 *In the following, the assertion holds.*

```c
def f() {
    int a;
    return &a;
}

def g() {
    int* p = f();
    // @ assert \texttt{\dangling\{Here\}(\&p)};
    return p+1;
}
```

In most cases, the arguments to \texttt{\dangling} are pointers to l-values that themselves have type pointer, so the usual signature of \texttt{\dangling} is actually \texttt{set<\alpha**>} \rightarrow bool. The signature \texttt{set<\alpha*> \rightarrow bool} is useful to handle pointer values that have been written inside scalar variables through heterogeneous casts.
Chapter 3
Libraries

Disclaimer: this chapter is unfinished, it is left here to give an idea of what it will look like in the final document.

This chapter is devoted to libraries of specification, built upon the ACSL specification language. Section 3.2 describes additional predicates introduced by the Jessie plugin of Frama-C, to propose a slightly higher level of annotation.

3.1 Libraries of logic specifications

A standard library is provided, in the spirit of the List module of Section 2.6.11

3.1.1 Real numbers

A library of general purpose functions and predicate over real numbers, floats and doubles.
Includes

- abs, exp, power, log, sin, cos, atan, etc. over reals
- isFinite predicate over floats and doubles (means not NaN nor infinity)
- rounding reals to floats or doubles with specific rounding modes.

3.1.2 Finite lists

- pure functions nil, cons, append, fold, etc.
- Path, Reachable, isFiniteList, isCyclic, etc. on C linked-lists.

3.1.3 Sets and Maps

Finite sets, finite maps, in ZB-style.
3.2 Jessie library: logical addressing of memory blocks

The Jessie library is a collection of logic specifications whose semantics is well-defined only on source codes free from architecture-dependent features. In particular it is currently incompatible with pointer casts or unions (although there is ongoing work to support some of them [20]). As a consequence, a valid pointer of some type $\tau$ necessarily points to a memory block which contains values of type $\tau$.

### 3.2.1 Abstract level of pointer validity

In the particular setting described above, it is possible to introduce the following logic functions:

- $\text{offset}_{\min}(p)$ is the minimum integer $i$ such that $(p+i)$ is a valid pointer.
- $\text{offset}_{\max}(p)$ is the maximum integer $i$ such that $(p+i)$ is a valid pointer.

The following properties hold:

- $\text{offset}_{\min}(p+i) = \text{offset}_{\min}(p) - i$
- $\text{offset}_{\max}(p+i) = \text{offset}_{\max}(p) - i$

It also introduces some syntactic sugar:

- $\text{valid}_{\min}(L)(p)$
- $\text{valid}_{\max}(L)(p)$

and the ACSL built-in predicate $\text{valid}(L)(p+(a..b))$ is now equivalent to $\text{valid}_{\min}(L)(p,a,b)$.

### 3.2.2 Strings

**Experimental**. The logic function

```
/*@ logic integer strlen(char* p);
*/
```

denotes the length of a 0-terminated C string. It is a total function, whose value is nonnegative if and only if the pointer in argument is really a string.

**Example 3.1** Here is a contract for the `strcpy` function:
3.3 Memory leaks

**Experimental**

Verification of absence of memory leak is outside the scope of the specification language. On the other hand, various models could be set up, using for example ghost variables.
This document presents a Behavioral Interface Specification Language for ANSI C source code. It provides a common basis that could be shared among several tools. The specification language described here is intended to evolve in the future, and remain open to additional constructions. One interesting possible extension regards “temporal” properties in a large sense, such as liveness properties, which can sometimes be simulated by regular specifications with ghost variables [11], or properties on evolution of data over the time, such as the history constraints of JML, or in the Lustre assertion language.
A.1 Glossary

pure expressions In ACSL setting, a pure expression is a C expression which contains no assignments, no incrementation operator ++ or --, no function call, and no access to a volatile object. The set of pure expression is a subset of the set of C expressions without side effect (C standard [13, 12], §5.1.2.3, alinea 2).

left-values A left-value (lvalue for short) is an expression which denotes some place in the memory during program execution, either on the stack, on the heap, or in the static data segment. It can be either a variable identifier or an expression of the form *e, e[e], e.id or e->id, where e is any expression and id a field name. See C standard, §6.3.2.1 for a more detailed description of lvalues.

A modifiable lvalue is an lvalue allowed in the left part of an assignment. In essence, all lvalues are modifiable except variables declared as const or of some array type with explicit length.

pre-state and post-state For a given function call, the pre-state denotes the program state at the beginning of the call, including the current values for the function parameters. The post-state denotes the program state at the return of the call.

function behavior A function behavior (behavior for short) is a set of properties relating the pre-state and the post-state for a possibly restricted set of pre-states (behavior assumptions).

function contract A function contract (contract for short) forms a specification of a function, consisting of the combination of a precondition (a requirement on the pre-state for any caller to that function), a collection of behaviors, and possibly a measure in case of a recursive function.

A.2 Comparison with JML

Although we took our inspiration in the Java Modeling Language (aka JML [15]), ACSL is notably different from JML in two crucial aspects:

- ACSL is a BISL for C, a low-level structured language, while JML is a BISL for Java, an object-oriented inheritance-based high-level language. Not only the language features
are not the same but the programming styles and idioms are very different, which entails also different ways of specifying behaviors. In particular, C has no inheritance nor exceptions, and no language support for the simplest properties on memory (e.g., the size of an allocated memory block).

- JML relies on runtime assertion checking (RAC) when typing, static analysis and automatic deductive verification fail. The example of CCured [21, 7], that adds strong typing to C by relying on RAC too, shows that it is not possible to do it in a modular way. Indeed, it is necessary to modify the layout of C data structures for RAC, which is not modular. The follow-up project Deputy [8] thus reduces the checking power of annotations in order to preserve modularity. On the contrary, we choose not to restrain the power of annotations (e.g., all first order logic formulas are allowed). To that end, we rely on manual deductive verification using an interactive theorem prover (e.g., Coq) when every other technique failed.

In the remainder of this chapter, we describe these differences in further details.

A.2.1 Low-level language vs. inheritance-based one

No inherited specifications

JML has a core notion of inheritance of specifications, that duplicates in specifications the inheritance feature of Java. Inheritance combined with visibility and modularity account for a number of complex features in JML (e.g., spec_public modifier, data groups, represents clauses, etc), that are necessary to express the desired inheritance-related specifications while respecting visibility and modularity. Since C has no inheritance, these intricacies are avoided in ACSL.

Error handling without exceptions

The usual way of signaling errors in Java is through exceptions. Therefore, JML specifications are tailored to express exceptional postconditions, depending on the exception raised. Since C has no exceptions, ACSL does not use exceptional specifications. Instead, C programmers are used to signal errors by returning special values, like mandated in various ways in the C standard.

Example A.1 In §7.12.1 of the standard, it is said that functions in <math.h> signal errors as follows: “On a domain error, [...] the integer expression errno acquires the value EDOM.”

Example A.2 In §7.19.5.1 of the standard, it is said that function fclose signals errors as follows: “The fclose function returns [...] EOF if any errors were detected.”

Example A.3 In §7.19.6.1 of the standard, it is said that function fprintf signals errors as follows: “The fprintf function returns [...] a negative value if an output or encoding error occurred.”

Example A.4 In §7.20.3 of the standard, it is said that memory management functions signal errors as follows: “If the space cannot be allocated, a null pointer is returned.”
A.2. COMPARISON WITH JML

As shown by these few examples, there is no unique way to signal errors in the C standard library, not mentioning user-defined functions. But since errors are signaled by returning special values, it is sufficient to write an appropriate postcondition:

```c
/*@ ensures \result == error_value || normal_postcondition; */
```

C contracts are not Java ones

In Java, the precondition of the following function that nullifies an array of characters is always true. Even if there was a precondition on the length of array \( a \), it could easily be expressed using the Java expression \( a.length \) that gives the dynamic length of array \( a \).

```java
public static void Java_nullify(char[] a) {
    if (a == null) return;
    for (int i = 0; i < a.length; ++i) {
        a[i] = 0;
    }
}
```

On the contrary, the precondition of the same function in C, whose definition follows, is more involved. First, remark that the C programmer has to add an extra argument for the size of the array, or rather a lower bound on this array size.

```c
void C_nullify(char* a, unsigned int n) {
    int i;
    if (n == 0) return;
    for (i = 0; i < n; ++i) {
        a[i] = 0;
    }
}
```

A correct precondition for this function is the following:

```c
/*@ requires \valid (a + 0..(n-1)); */
```

where predicate \( \valid \) is the one defined in Section 2.7.1. (note that \( \valid (a + 0..(-1)) \) is the same as \( \valid (\emptyset) \) and thus is true regardless of the validity of \( a \) itself). When \( n \) is null, \( a \) does not need to be valid at all, and when \( n \) is strictly positive, \( a \) must point to an array of size at least \( n \). To make it more obvious, the C programmer adopted a defensive programming style, which returns immediately when \( n \) is null. We can duplicate this in the specification:

```c
/*@ requires n == 0 || \valid(a + 0..(n-1)); */
```

Usually, many memory requirements are only necessary for some paths through the function, which correspond to some particular behaviors, selected according to some tests performed along the corresponding paths. Since C has no memory primitives, these tests involve other variables that the C programmer added to track additional information, like \( n \) in our example.

To make it easier, it is possible in ACSL to distinguish between the \texttt{assumes} part of a behavior, that specifies the tests that need to succeed for this behavior to apply, and the \texttt{requires} part that specifies the additional preconditions that must be true when a behavior applies. The specification for our example can then be translated into:

```c
/*@ behavior n_is_null:
  @ assumes n == 0;
*/

/*@ behavior n_is_not_null:
  @ requires n == 0 || \valid(a + 0..(n-1)); */
```
This is equivalent to the previous requirement, except here behaviors can be completed with postconditions that belong to one behavior only. Contrary to JML, the set of behaviors for a function do not necessarily cover all cases of use for this function, as mentioned in Section 2.3.3. This allows for partial specifications, whereas JML behaviors cannot offer such flexibility. Here, our two behaviors are clearly mutually exclusive, and, since n is an unsigned int, our they cover all the possible cases. We could have specified that as well, by adding the following lines in the contract (see Section 2.3.3).

ACSL contracts vs. JML ones
To fully understand the difference between specifications in ACSL and JML, we detail in below the requirement on the pre-state and the guarantee on the post-state given by behaviors in JML and ACSL.

A JML contract is either *lightweight* or *heavyweight*. For the purpose of our comparison, it is sufficient to know that a lightweight contract has *requires* and *ensures* clauses all at the same level, while an heavyweight contract has multiple behaviors, each consisting of *requires* and *ensures* clauses. Although it is not possible in JML to mix both styles, we can define here what it would mean to have both, by conjoining the conditions on the pre- and the post-state. Here is an hypothetical JML contract mixing lightweight and heavyweight styles:

```
@* @ requires P1;
@ requires P2;
@ ensures Q1;
@ ensures Q2;
@ behavior x1:
@ requires A1;
@ requires R1;
@ ensures E1;
@ behavior x2:
@ requires A2;
@ requires R2;
@ ensures E2;
@*/
```

It assumes from the pre-state the condition:

\[
P_1 \land P_2 \land (A_1 \land R_1) \lor (A_2 \land R_2)
\]

and guarantees that the following condition holds in post-state:

\[
Q_1 \land Q_2 \land (\text{old}(A_1 \land R_1) \Rightarrow E_1) \land (\text{old}(A_2 \land R_2) \Rightarrow E_2)
\]

Here is now an ACSL specification:

```
@* requires P1;
@ requires P2;
```
A.2. COMPARISON WITH JML

Syntactically, the only difference with the JML specification is the addition of the assumes clauses. Its translation to assume-guarantee is however quite different. It assumes from the pre-state the condition:

\[ P_1 \land P_2 \land (A_1 \implies R_1) \land (A_2 \implies R_2) \]

and guarantees that the following condition holds in the post-state:

\[ Q_1 \land Q_2 \land (\neg \text{old}(A_1) \implies E_1) \land (\neg \text{old}(A_2) \implies E_2) \]

Thus, ACSL allows to distinguish between the clauses that control which behavior is active (the assumes clauses) and the clauses that are preconditions for a particular behavior (the internal requires clauses). In addition, as mentioned above, there is by default no requirement in ACSL for the specification to be complete (The last part of the JML condition on the pre-state). If desired, this has to be precised explicitly with a complete behaviors clause as seen in Section 2.3.3.

A.2.2 Deductive verification vs. RAC

Sugar-free behaviors

As explained in details in [22], JML heavyweight behaviors can be viewed as syntactic sugar (however complex it is) that can be translated automatically into more basic contracts consisting mostly of pre- and postconditions and frame conditions. This allows complex nesting of behaviors from the user point of view, while tools only have to deal with basic contracts. In particular, the major tools on JML use this desugaring process, like the Common JML tools to do assertion checking, unit testing, etc. (see [18]) and the tool ESC/Java2 for automatic deductive verification of JML specifications (see [6]).

One issue with such a desugaring approach is the complexity of the transformations involved, as e.g. for desugaring assignable clauses between multiple spec-cases in JML [22]. Another issue is precisely that tools only see one global contract, instead of multiple independent behaviors, that could be analyzed separately in more detail. Instead, we favor the view that a function implements multiple behaviors, that can be analyzed separately if a tool feels like it. Therefore, we do not intend to provide a desugaring process.

Axiomatized functions in specifications

JML only allows pure Java methods to be called in specifications [16]. This is certainly understandable when relying on RAC: methods called should be defined so that the runtime can call them, and they should not have side-effects in order not to pollute the program they are supposed to annotate.
In our setting, it is desirable to allow calls to logical functions in specifications. These functions may be defined, like program ones, but they may also be only declared (with a suitable declaration of `reads` clause) and defined through an axiomatization. This makes for richer specifications that may be useful either in automatic or in manual deductive verification.

### A.2.3 Syntactic differences

The following table summarizes the difference between JML keywords and ACSL ones, when the intent is the same, although minor differences might exist.

<table>
<thead>
<tr>
<th>JML</th>
<th>ACSL</th>
</tr>
</thead>
<tbody>
<tr>
<td>modifiable,assignable</td>
<td>assigns</td>
</tr>
<tr>
<td>measured_by</td>
<td>decreases</td>
</tr>
<tr>
<td>loop_invariant</td>
<td>loop invariant</td>
</tr>
<tr>
<td>decreases</td>
<td>loop variant</td>
</tr>
<tr>
<td>( (\forall \tau x ; P ; Q) )</td>
<td>( (\forall \tau x ; P \Rightarrow Q) )</td>
</tr>
<tr>
<td>( (\exists \tau x ; P ; Q) )</td>
<td>( (\exists \tau x ; P \land Q) )</td>
</tr>
<tr>
<td>( \max \tau x ; a &lt;= x &lt;= b ; f )</td>
<td>( \max(a,b,\lambda \tau x ; f) )</td>
</tr>
</tbody>
</table>

### A.3 Typing rules

Disclaimer: this section is unfinished, it is left here just to give an idea of what it will look like in the final document.

#### A.3.1 Rules for terms

Integer promotion:

\[ \frac{}{\Gamma \vdash e : \tau} \quad \frac{}{\Gamma \vdash e : \text{integer}} \]

if \( \tau \) is any C integer type `char`, `short`, `int`, or `long`, whatever attribute they have, in particular signed or unsigned

Variables:

\[ \frac{}{\Gamma \vdash id : \tau} \quad \text{if } id : \tau \in \Gamma \]

Unary integer operations:

\[ \frac{}{\Gamma \vdash t : \text{integer}} \quad \frac{}{\Gamma \vdash \text{op } t : \text{integer}} \quad \text{if } op \in \{+, -, \sim\} \]

Boolean negation:

\[ \frac{}{\Gamma \vdash t : \text{boolean}} \quad \frac{}{\Gamma \vdash \neg t : \text{boolean}} \]

Pointer dereferencing:

\[ \frac{}{\Gamma \vdash t : \tau^*} \quad \frac{}{\Gamma \vdash *t : \tau} \]

Address operator:

\[ \frac{}{\Gamma \vdash t : \tau} \quad \frac{}{\Gamma \vdash \&t : \tau^*} \]
A.4. SPECIFICATION TEMPLATES

Binary

\[
\begin{align*}
\Gamma & \vdash t_1 : \text{integer} & \Gamma & \vdash t_2 : \text{integer} \\
\Gamma & \vdash t_1 \ op \ t_2 : \text{integer} & \text{if } op & \in \{+,-,\ast,/,\%\} \\
\Gamma & \vdash t_1 : \text{real} & \Gamma & \vdash t_2 : \text{real} \\
\Gamma & \vdash t_1 \ op \ t_2 : \text{real} & \text{if } op & \in \{+,-,\ast,/,\}\end{align*}
\]

\[
\begin{align*}
\Gamma & \vdash t_1 : \text{integer} & \Gamma & \vdash t_2 : \text{integer} \\
\Gamma & \vdash t_1 \ op \ t_2 : \text{boolean} & \text{if } op & \in \{==!,==,<,<=,>,>=\} \\
\Gamma & \vdash t_1 : \text{real} & \Gamma & \vdash t_2 : \text{real} \\
\Gamma & \vdash t_1 \ op \ t_2 : \text{real} & \text{if } op & \in \{==!,==,<,<=,>,>=\} \\
\Gamma & \vdash t_1 : \tau & \Gamma & \vdash t_2 : \tau \\
\Gamma & \vdash t_1 \ op \ t_2 : \text{boolean} & \text{if } op & \in \{==!,==,<,<=,>,>=\}
\end{align*}
\]

(to be continued)

A.3.2 Typing rules for sets

We consider the typing judgement \( \Gamma, \Lambda \vdash s : \tau, b \) meaning that \( s \) is a set of terms of type \( \tau \), which is moreover a set of locations if the boolean \( b \) is true. \( \Gamma \) is the C environment and \( \Lambda \) is the logic environment.

Rules:

\[
\begin{align*}
\Gamma, \Lambda & \vdash id : \tau, \text{true} & \text{if } id & \in \Gamma \\
\Gamma, \Lambda & \vdash id : \tau, \text{true} & \text{if } id & \in \Lambda \\
\Gamma, \Lambda & \vdash s : \tau*, b \\
\Gamma, \Lambda & \vdash *s : \tau, \text{true} \\
\Gamma & \vdash s \rightarrow id : set < \text{struct} S > \\
\Gamma, \Lambda & \vdash b \cup \Lambda \vdash e : tset\tau \\
\Gamma, \Lambda & \vdash \{ e \mid \ b; \ P \} : tset\tau \\
\Gamma, \Lambda & \vdash e_1 : \tau, b & \Gamma, \Lambda & \vdash e_2 : \tau, b \\
\Gamma, \Lambda & \vdash e_1, e_2 : \tau, b
\end{align*}
\]

A.4 Specification Templates

This section describes some common issues that may occur when writing an ACSL specification and proposes some solution to overcome them.

A.4.1 Accessing a C variable that is masked

The situation may happen where it is necessary to refer in an annotation to a C variable that is masked at that point. For instance, a function contract may need to refer to a global variable that has the same name as a function parameter, as in the following code:
In order to write the `assigns` clause for `f`, we must access the global variable `x`, since `f` calls `g`, which can modify `x`. This is not possible with C scoping rules, as `x` refers to the parameter of `f` in the scope of the function.

A solution is to use a ghost pointer to `x`, as shown in the following code:

```cpp
int x;
//@ assigns x;
int g();
//@ assigns x;
int f(int x) {
  // ...
  return g();
}
```

A.5 Illustrative example

This is an attempt to define an example for ACSL, much as the Purse example in JML description papers. It is a memory allocator, whose main functions are `memory_alloc` and `memory_free`, to respectively allocate and deallocate memory. The goal is to exercise as much as possible of ACSL.

```cpp
#include <stdlib.h>
define DEFAULT_BLOCK_SIZE 1000
typedef enum _bool { false = 0, true = 1 } bool;
//@ predicate finite_list<A>((A* -> A*) next_elem, A* ptr) =
  @ ptr == \null ||
  @ (\valid(ptr) && finite_list(next_elem,next_elem(ptr)));
//@ logic integer list_length<A>((A* -> A*) next_elem, A* ptr) =
  @ (ptr == \null) ? 0 :
  @ 1 + list_length(next_elem,next_elem(ptr));
//@ predicate lower_length<A>((A* -> A*) next_elem, A* ptr1, A* ptr2) =
  @ finite_list(next_elem, ptr1) && finite_list(next_elem, ptr2)
  @ && list_length(next_elem, ptr1) < list_length(next_elem, ptr2);
A.5. ILLUSTRATIVE EXAMPLE

21 /**
22     // forward reference
23 typedef struct _memory_slice;
24 */
25
26 /* A memory block holds a pointer to a raw block of memory allocated by
27       * calling [malloc]. It is sliced into chunks, which are maintained by
28       * the [slice] structure. It maintains additional information such as
29       * the [size] of the memory block, the number of bytes [used] and the [next]
30       * index at which to put a chunk.
31 */
32 typedef struct _memory_block {
33     //@
34     // ghost field [packed] is meant to be used as a guard that tells when
35     // the invariant of a structure of type [memory_block] holds
36     unsigned int size;
37     // size of the array [data]
38     unsigned int next;
39     // next index in [data] at which to put a chunk
40     unsigned int used;
41     // how many bytes are used in [data], not necessarily contiguous ones
42     char* data;
43     // raw memory block allocated by [malloc]
44     struct _memory_slice* slice;
45     // structure that describes the slicing of a block into chunks
46 } memory_block;
47
48 /*@
49     strong type invariant
50     inv_memory_block(memory_block mb) =
51       (@ mb.packed ==>
52         (0 < mb.size && mb.used <= mb.next <= mb.size
53          && \offset(mb.data) == 0
54          && \block_length(mb.data) == mb.size)
55          && valid_memory_block(mb)) && mb->packed ;
56     @*/
57
58 /* A memory chunk holds a pointer [data] to some part of a memory block
59       * [block]. It maintains the [offset] at which it points in the block, as well
60       * as the [size] of the block it is allowed to access. A field [free] tells
61       * whether the chunk is used or not.
62 */
63 typedef struct _memory_chunk {
64     //@
65     // ghost field [packed] is meant to be used as a guard that tells when
66     // the invariant of a structure of type [memory_chunk] holds
67     unsigned int offset;
68     // offset at which [data] points into [block->data]
69     unsigned int size;
70     // size of the chunk
71     bool free;
72     // true if the chunk is not used, false otherwise
73     memory_block* block;
74     // block of memory into which the chunk points
75     char* data;
76     // shortcut for [block->data + offset]
APPENDIX A. APPENDICES

```c
} memory_chunk;

/*@ strong type invariant inv_memory_chunk(memory_chunk mc) =
   @ mc.packed ==> 
   @ (0 < mc.size && valid_memory_block(mc.block) 
   @ && mc.offset + mc.size <= mc.block->next) ;
   @
   @ predicate valid_memory_chunk(memory_chunk* mc, int s) = 
   @ \valid(mc) && mc->packed && mc->size == s ;
   @
   @ predicate used_memory_chunk(memory_chunk mc) = 
   @ mc.free == false ;
   @
   @ predicate freed_memory_chunk(memory_chunk mc) = 
   @ mc.free == true ;
   @*/

/*@ A memory chunk list links memory chunks in the same memory block.
   * Newly allocated chunks are put first, so that the offset of chunks
   * decreases when following the [next] pointer. Allocated chunks should
   * fill the memory block up to its own [next] index.
   */
typedef struct _memory_chunk_list {
  memory_chunk* chunk;
  // current list element
  struct _memory_chunk_list* next;
  // tail of the list
} memory_chunk_list;

/*@ logic memory_chunk_list* next_chunk(memory_chunk_list* ptr) =
   @ ptr->next ;
   @
   @ predicate valid_memory_chunk_list
   @ (memory_chunk_list* mcl, memory_block* mb) = 
   @ \valid(mcl) && valid_memory_chunk(mcl->chunk,mcl->chunk->size) 
   @ && mcl->chunk->block == mb 
   @ && (mcl->next == \null ||
   @ valid_memory_chunk_list(mcl->next, mb)) 
   @ && mcl->offset == mcl->chunk->offset 
   @ && (mcl->next == \null && mcl->chunk->offset == 0) ||
   @ (mcl->next != \null 
   @ && mcl->next->chunk->offset + mcl->next->chunk->size 
   @ == mcl->chunk->offset) 
   @)
   @ && finite_list(next_chunk, mcl) ;
   @
   @ predicate valid_complete_chunk_list
   @ (memory_chunk_list* mcl, memory_block* mb) = 
   @ valid_memory_chunk_list(mcl,mb) 
   @ && mcl->next->chunk->offset +
   @ mcl->next->chunk->size == mb->next ;
```
A.5. ILLUSTRATIVE EXAMPLE

/* A memory slice holds together a memory block [block] and a list of chunks
 * [chunks] on this memory block. */

typedef struct _memory_slice {
    // ghost field [packed] is meant to be used as a guard that tells when
    // the invariant of a structure of type [memory_slice] holds
    memory_block* block;
    memory_chunk_list* chunks;
} memory_slice;

/*@ strong type invariant inv_memory_slice(memory_slice* ms) =
...*/

/*@ A memory slice list links memory slices, to form a memory pool. */

typedef struct _memory_slice_list {
    // ghost field [packed] is meant to be used as a guard that tells when
    // the invariant of a structure of type [memory_slice_list] holds
    memory_slice* slice;
    struct _memory_slice_list* next;
} memory_slice_list;

/*@ logic memory_slice_list* next_slice(memory_slice_list* ptr) =
...*/

/*@ strong type invariant inv_memory_slice_list(memory_slice_list* msl) =
...*/
typedef memory_slice_list* memory_pool;

/*@ type invariant valid_memory_pool(memory_pool *mp) =
@  \valid(mp) && valid_memory_slice_list(*mp) ;
@*/

/*@ behavior zero_size:
@  assumes s == 0;
@  assigns \nothing;
@  ensures \result == 0;
@*/

/*@ behavior positive_size:
@  assumes s > 0;
@  requires valid_memory_pool(arena);
@  ensures \result == 0
@ || (valid_memory_chunk(\result,s) &&
@    used_memory_chunk(*\result));
@*/

memory_chunk* memory_alloc(memory_pool* arena, unsigned int s) {
  memory_slice_list *msl = *arena;
  memory_chunk_list *mcl;
  memory_slice *ms;
  memory_block *mb;
  memory_chunk *mc;
  unsigned int mb_size;
 //@
  ghost unsigned int mcl_offset;
  char *mb_data;
  // guard condition
  if (s == 0) return 0;
  // iterate through memory blocks (or slices)
 /*@*/
  while (msl != 0) {
    ms = msl->slice;
    mb = ms->block;
    mcl = ms->chunks;
    // does [mb] contain enough free space?
    if (s <= mb->size - mb->next) {
      //@ ghost ms->ghost = false; // unpack the slice
      // allocate a new chunk
      mc = (memory_chunk*)malloc(sizeof(memory_chunk));
      if (mc == 0) return 0;
      mc->offset = mb->next;
      mc->size = s;
      mc->free = false;
      mc->block = mb;
      //@ ghost mc->ghost = true; // pack the chunk
      // update block accordingly
      //@ ghost mb->ghost = false; // unpack the block
      mb->next += s;
      mb->used += s;
      //@ ghost mb->ghost = true; // pack the block
      // add the new chunk to the list
      mcl = (memory_chunk_list*)malloc(sizeof(memory_chunk_list));
if (mcl == 0) return 0;
mcl->chunk = mc;
mcl->next = ms->chunks;
ms->chunks = mcl;
//@ ghost ms->ghost = true; // pack the slice
return mc;
}

// iterate through memory chunks
/*@ 
@ loop invariant valid_memory_chunk_list(mcl,mb);
@ loop variant mcl for chunk_lower_length;
@ */
while (mcl != 0) {
  mc = mcl->chunk;
  // is [mc] free and large enough?
  if (mc->free && s <= mc->size) {
    mc->free = false;
    mb->used += mc->size;
    return mc;
  }
  // try next chunk
  mcl = mcl->next;
}

// allocate a new block
mb_size = (DEFAULT_BLOCK_SIZE < s) ? s : DEFAULT_BLOCK_SIZE;
mb_data = (char*)malloc(mb_size);
if (mb_data == 0) return 0;
mb = (memory_block*)malloc(sizeof(memory_block));
if (mb == 0) return 0;
mb->size = mb_size;
mb->next = s;
mb->used = s;
mb->data = mb_data;
//@ ghost mb->ghost = true; // pack the block
//@ allocate a new chunk
mc = (memory_chunk*)malloc(sizeof(memory_chunk));
if (mc == 0) return 0;
mc->offset = 0;
mc->size = s;
mc->free = false;
mc->block = mb;
//@ ghost mc->ghost = true; // pack the chunk
//@ allocate a new chunk list
mcl = (memory_chunk_list*)malloc(sizeof(memory_chunk_list));
if (mcl == 0) return 0;
//@ ghost mcl->offset = 0;
mcl->chunk = mc;
mcl->next = 0;
//@ allocate a new slice
ms = (memory_slice*)malloc(sizeof(memory_slice));
if (ms == 0) return 0;
ms->block = mb;
ms->chunks = mcl;
//@ ghost ms->ghost = true; // pack the slice
// update the block accordingly
mb->slice = ms;
// add the new slice to the list
msl = (memory_slice_list*)malloc(sizeof(memory_slice_list));
if (msl == 0) return 0;
msl->slice = ms;
msl->next = *arena;
//@ ghost msl->ghost = true; // pack the slice list
*arena = msl;
return mc;
}

/** behavior null_chunk:
 * assumes chunk == \null;
 * assigns \nothing;
 * @
 * behavior valid_chunk:
 * assumes chunk != \null;
 * requires valid_memory_pool(arena);
 * requires valid_memory_chunk(chunk,chunk->size);
 * requires used_memory_chunk(chunk);
 * ensures
 *   // if it is not the last chunk in the block, mark it as free
 *   (valid_memory_chunk(chunk,chunk->size)
 *    && freed_memory_chunk(chunk))
 *   ||
 *   // if it is the last chunk in the block, deallocate the block
 *   ! valid(chunk);
 * @*/
void memory_free(memory_pool* arena, memory_chunk* chunk) {
memory_slice_list *msl = *arena;
memory_block *mb = chunk->block;
memory_slice *ms = mb->slice;
memory_chunk_list *mcl;
memory_chunk *mc;

// is it the last chunk in use in the block?
if (mb->used == chunk->size) {
   // remove the corresponding slice from the memory pool
   // case it is the first slice
   if (msl->slice == ms) {
      *arena = msl->next;
      //@ ghost msl->ghost = false; // unpack the slice list
      free(msl);
   }
   // case it is not the first slice
   while (msl != 0) {
      if (msl->next != 0 && msl->next->slice == ms) {
         memory_slice_list* msl_next = msl->next;
         msl->next = msl->next->next;
         // unpack the slice list
         //@ ghost msl_next->ghost = false;
         free(msl_next);
         break;
      }
      msl = msl->next;
   }
}
A.6 Changes

A.6.1 Version 1.9

- Fix typo in definition of `\fresh` predicate (section 2.7.3)
- Fix grammar inconsistencies
  - use proper C rules names
  - fix mismatch in non-terminal names
- Rename 'Unspecified values' to "Dangling pointers" and precise it. (section 2.13.2)

A.6.2 Version 1.8

- Mention binary literal constant typing
A.6.3 Version 1.7

- Added missing shift operators in figure 2.1
- Modified syntax for naming terms and predicates (figures 2.2 and 2.1)
- Added syntax rule for literal constants (figure 2.1)

A.6.4 Version 1.6

- Modified syntax for model fields (section 2.11.2)
- Added missing logical xor operator (figure 2.1).
- Addition of logical labels related to loops (section 2.4.3).
- Addition of labels to built-ins related to memory blocks (section 2.7.1)
- Introduction of \valid\_read built-in and clarification of the notion of validity (section 2.7.1).
- Introduction of \allocable, \allocation, \freeable and \fresh (section 2.7.3).
- Introduction of allocates and frees clauses (section 2.7.3).
- Clarify the semantics of assigns clauses into statement contract.
- Improvements to the volatile clause (section 2.12.1).
- Clarify the evaluation of arrays inside an at (section 2.4.3).

A.6.5 Version 1.5

- Clarify the status of loop invariant in presence of break or side-effects in the loop test.
- Introduction of \with keyword for functional updates.
- Added bnf entry for completeness of function behaviors.
- Order of clauses in statement contracts is now fixed.
- requires clauses are allowed before behaviors of statement contracts.
- Added explicit singleton construct for sets.
- Introduction of logical arrays.
- Operations over pointers and arrays have been precised.
- Predicate \initialized (section 2.13.1) now takes a set of pointers as argument.
A.6. CHANGES

A.6.6 Version 1.4

- Added UTF-8 counterparts for built-in types (integer, real, boolean).
- Fixed typos in the examples corresponding to features implemented in Frama-C.
- Order of clauses in function contracts is now fixed.
- Introduction of abrupt termination clauses.
- Introduction of axiomatic to gather predicates, logic functions, and their defining axioms.
- Added specification templates appendix for common specification issues.
- Use of sets as first-class term has been precised.
- Fixed semantics of predicate \textbackslash separated.

A.6.7 Version 1.3

- Functional update of structures.
- Terminates clause in function behaviors.
- Typos reported by David Mentré.

A.6.8 Version 1.2

This is the first public release of this document.
Bibliography


http://kind.ucd.ie/products/opensource/ESCJava2/.

http://www.eecs.ucf.edu/~leavens/JML/.


List of Figures

2.1 Grammar of terms ........................................... 16
2.2 Grammar of predicates ..................................... 17
2.3 Grammar of binders and type expressions ............... 18
2.4 Operator precedence ......................................... 19
2.5 Grammar of function contracts ............................. 29
2.6 \texttt{\textbackslash old} and \texttt{\textbackslash result} in terms ................................................. 29
2.7 Grammar for sets of terms ................................. 34
2.8 Grammar for assertions ................................. 36
2.9 Grammar for loop annotations ............................. 37
2.10 Grammar for general inductive invariants ............... 39
2.11 Grammar for statement contracts ......................... 44
2.12 Grammar for global logic definitions .................... 48
2.13 Grammar for inductive definitions ........................ 49
2.14 Grammar for axiomatic declarations ...................... 50
2.15 Grammar for higher-order constructs .................... 52
2.16 Grammar for concrete logic types and pattern-matching ................................................. 53
2.17 Grammar for logic declarations with labels .............. 54
2.18 Grammar for logic declarations with \texttt{reads} clauses ................................................. 56
2.19 Grammar extension of terms and predicates about memory ................................................. 57
2.20 Grammar for dynamic allocations and deallocations ... 59
2.21 Grammar of contracts about abrupt terminations ........ 64
2.22 Grammar for dependencies information ............... 66
2.23 Grammar for declarations of data invariants .......... 68
2.24 Grammar for declarations of model variables and fields ................................................. 70
2.25 Grammar for ghost statements ......................... 73
2.26 Grammar for volatile constructs ......................... 75
disjoint behaviors, 33
do, 37
\dynamic, 59
dynamic allocation, 59
else, 73
\empty, 34
ensures, 29, 30, 44, 64
\eq_double, 23
\eq_float, 23
\exists, 17
\exit_status, 64, 65
exits, 64
\false, 16, 17, 20
for, 29, 36, 37, 39, 44
\forall, 17
formal parameter, 29, 30, 42, 68
\freeable, 57, 60
frees, 59, 66
\fresh, 57, 60
\from, 66
function behavior, 31, 85
function contract, 28, 85
functional expression, 66
\ge_double, 23
\ge_float, 23
ghost, 72, 73
global, 68
global invariant, 67
grammar entries
C-compound-statement, 36
C-external-declaration, 48
C-statement, 36
abrupt-clause-fn, 64
abrupt-clause-stmt, 64
allocation-clause, 59
assertion, 36, 39
assigns-clause, 29, 66
assumes-clause, 29
axiom-decl, 50

? , 16, 17
_ , 53
abrupt clause, 64
\allocable, 57, 60
allocates, 59, 66
allocation, 59
\allocation, 57, 60
allocation_status, 59
annotation, 36
  loop, 36
as, 53
assert, 36
assertion, 36
assigns, 29, 30, 37, 44, 66
assumes, 29
\at, 41
\automatic, 59
axiom, 50
axiomatic, 50
\base_addr, 57, 58
behavior, 29, 31, 44, 85
behaviors, 29
\block_length, 28, 57, 58
boolean, 18, 20
breaks, 64, 65
case, 49, 53
cast, 20–23
complete, 29
complete behaviors, 33
comprehension, 35
continues, 64, 65
contract, 28, 36, 43, 85
\danling, 77
data invariant, 67
deallocation, 59
decreases, 29, 45
dependency, 66
disjoint, 29
<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>axiomatic-decl</td>
<td>50</td>
</tr>
<tr>
<td>behavior-body-stmt</td>
<td>44</td>
</tr>
<tr>
<td>behavior-body</td>
<td>29</td>
</tr>
<tr>
<td>bin-op</td>
<td>16</td>
</tr>
<tr>
<td>binders</td>
<td>18</td>
</tr>
<tr>
<td>binder</td>
<td>18</td>
</tr>
<tr>
<td>breaks-clause</td>
<td>64</td>
</tr>
<tr>
<td>built-in-logic-type</td>
<td>18</td>
</tr>
<tr>
<td>completeness-clause</td>
<td>29</td>
</tr>
<tr>
<td>constructor</td>
<td>53</td>
</tr>
<tr>
<td>continues-clause</td>
<td>64</td>
</tr>
<tr>
<td>data-inv-decl</td>
<td>68</td>
</tr>
<tr>
<td>data-invariant</td>
<td>68</td>
</tr>
<tr>
<td>declaration</td>
<td>68</td>
</tr>
<tr>
<td>decreases-clause</td>
<td>29</td>
</tr>
<tr>
<td>direct-declarator</td>
<td>73</td>
</tr>
<tr>
<td>dyn-allocation-addresses</td>
<td>59</td>
</tr>
<tr>
<td>ensures-clause</td>
<td>29</td>
</tr>
<tr>
<td>exits-clause</td>
<td>64</td>
</tr>
<tr>
<td>ext-quantifier</td>
<td>52</td>
</tr>
<tr>
<td>function-contract</td>
<td>29</td>
</tr>
<tr>
<td>ghost-selection-statement</td>
<td>73</td>
</tr>
<tr>
<td>ghost-type-specifier</td>
<td>73</td>
</tr>
<tr>
<td>inductive-def</td>
<td>49</td>
</tr>
<tr>
<td>inv-strength</td>
<td>68</td>
</tr>
<tr>
<td>label-binders</td>
<td>54</td>
</tr>
<tr>
<td>lemma-decl</td>
<td>48</td>
</tr>
<tr>
<td>literal</td>
<td>16</td>
</tr>
<tr>
<td>location-addresses</td>
<td>57</td>
</tr>
<tr>
<td>location-address</td>
<td>57</td>
</tr>
<tr>
<td>locations</td>
<td>29</td>
</tr>
<tr>
<td>location</td>
<td>29</td>
</tr>
<tr>
<td>logic-const-decl</td>
<td>50</td>
</tr>
<tr>
<td>logic-const-def</td>
<td>48</td>
</tr>
<tr>
<td>logic-decl</td>
<td>50</td>
</tr>
<tr>
<td>logic-def</td>
<td>48–50, 53</td>
</tr>
<tr>
<td>logic-function-decl</td>
<td>50, 56</td>
</tr>
<tr>
<td>logic-function-def</td>
<td>48, 56</td>
</tr>
<tr>
<td>logic-predicate-decl</td>
<td>50, 56</td>
</tr>
<tr>
<td>logic-predicate-def</td>
<td>48, 56</td>
</tr>
<tr>
<td>logic-type-decl</td>
<td>50</td>
</tr>
<tr>
<td>logic-type-def</td>
<td>53</td>
</tr>
<tr>
<td>logic-type-expr</td>
<td>18</td>
</tr>
<tr>
<td>logic-type</td>
<td>50</td>
</tr>
<tr>
<td>loop-allocation</td>
<td>59</td>
</tr>
<tr>
<td>loop-annot</td>
<td>37</td>
</tr>
<tr>
<td>loop-assigns</td>
<td>37</td>
</tr>
<tr>
<td>loop-behavior</td>
<td>37</td>
</tr>
<tr>
<td>loop-clause</td>
<td>37</td>
</tr>
<tr>
<td>loop-invariant</td>
<td>37</td>
</tr>
<tr>
<td>loop-variant</td>
<td>37</td>
</tr>
<tr>
<td>match-cases</td>
<td>37</td>
</tr>
<tr>
<td>match-case</td>
<td>53</td>
</tr>
<tr>
<td>named-behavior-stmt</td>
<td>44</td>
</tr>
<tr>
<td>named-behavior</td>
<td>29</td>
</tr>
<tr>
<td>one-label</td>
<td>57</td>
</tr>
<tr>
<td>parameters</td>
<td>48</td>
</tr>
<tr>
<td>parameter</td>
<td>48</td>
</tr>
<tr>
<td>pat</td>
<td>53</td>
</tr>
<tr>
<td>poly-id</td>
<td>48, 54</td>
</tr>
<tr>
<td>postfix-expression</td>
<td>73</td>
</tr>
<tr>
<td>pred</td>
<td>17, 29, 34, 57</td>
</tr>
<tr>
<td>reads-clause</td>
<td>56</td>
</tr>
<tr>
<td>record-type</td>
<td>53</td>
</tr>
<tr>
<td>rel-op</td>
<td>17</td>
</tr>
<tr>
<td>requires-clause</td>
<td>29</td>
</tr>
<tr>
<td>returns-clause</td>
<td>64</td>
</tr>
<tr>
<td>simple-clause-stmt</td>
<td>44</td>
</tr>
<tr>
<td>simple-clause</td>
<td>29</td>
</tr>
<tr>
<td>statement-contract</td>
<td>44</td>
</tr>
<tr>
<td>statements-ghost</td>
<td>73</td>
</tr>
<tr>
<td>statement</td>
<td>37, 44, 73</td>
</tr>
<tr>
<td>struct-declaration</td>
<td>73</td>
</tr>
<tr>
<td>sum-type</td>
<td>53</td>
</tr>
<tr>
<td>terminates-clause</td>
<td>29</td>
</tr>
<tr>
<td>term</td>
<td>16, 29, 52, 53, 57, 64</td>
</tr>
<tr>
<td>tset</td>
<td>34</td>
</tr>
<tr>
<td>two-labels</td>
<td>57</td>
</tr>
<tr>
<td>type-expr</td>
<td>18, 48, 53</td>
</tr>
<tr>
<td>type-invariant</td>
<td>68</td>
</tr>
<tr>
<td>type-var-binders</td>
<td>48</td>
</tr>
<tr>
<td>type-var</td>
<td>48</td>
</tr>
<tr>
<td>unary-op</td>
<td>16</td>
</tr>
<tr>
<td>variable-ident</td>
<td>18</td>
</tr>
</tbody>
</table>

\_\_double, 23
\_\_float, 23

Here, 41, 65

hybrid
  function, 54
  predicate, 54

if, 73
inductive, 49
inductive definitions, 49
inductive predicates, 49
\_\_initialized, 76
integer, 18, 20
INDEX

\inter, 34
invariant, 39
  data, 67
  global, 67
  loop, 37
  strong, 67
  type, 67
  weak, 67
invariant, 37, 39, 68
\is_finite, 24
\is_NaN, 24
l-value, 34
\lambda, 52
\le_double, 23
\le_float, 23
left-value, 85
lemma, 48
\let, 16, 17, 53
library, 57
location, 34, 63
logic, 48, 50, 56
logic specification, 47
loop
  allocation, 62
  annotation, 36
  assigns, 37
  behavior, 38
  deallocation, 62
  invariant, 37
  variant, 39
loop, 37, 59
LoopCurrent, 41
LoopEntry, 41
\lt_double, 23
\lt_float, 23
lvalue, 85
\match, 53
\max, 52
\min, 52
model, 70
module, 56
\ne_double, 23
\ne_float, 23
\nothing, 29
\null, 57, 58
\numof, 52
\offset, 57, 58
Old, 41, 65
\old, 29, 41, 65
polymorphism, 51
Post, 41, 65
post-state, 85
Pre, 41, 65
pre-state, 85
predicate, 16
predicate, 48, 50, 56
\product, 52
pure expression, 85
reads, 56, 75
real, 18, 20
real_of_double, 23
real_of_float, 23
recursion, 51
\register, 59
requires, 29, 30
\result, 29, 65
returns, 64, 65
\round_double, 23
\round_float, 23
\separated, 57, 59
set type, 63
sizeof, 16, 22
specification, 47
statement contract, 36, 43
\static, 59
strong, 68
\subset, 34
\sum, 52
term, 16
termination, 29, 47
termination, 39, 44
\true, 16, 17, 20
type
  concrete, 53
  polymorphic, 51
  record, 53
  sum, 53
type, 50, 53, 68
type invariant, 67
\unallocated, 59
\union, 34
\valid, 57, 58
\texttt{valid\_read}, 57, 58
\texttt{variant}, 37, 39, 45
\texttt{volatile}, 75

\texttt{weak}, 68
\texttt{while}, 37
\texttt{\_with}, 16
\texttt{writes}, 75