Preprocessor Specification

Grammar Reference

Copyright © 2020-2021, Hugo Décharnes. All rights reserved.

THE INFORMATION CONTAINED IN THIS DOCUMENT IS PROVIDED ON AN “AS IS” BASIS, AND IS SUBJECT TO CHANGE WITHOUT NOTICE. NOTHING IN THIS DOCUMENT SHALL OPERATE AS AN EXPRESS OR IMPLIED LICENSE OR INDEMNITY UNDER THE INTELLECTUAL PROPERTY RIGHTS. IN NO EVENT WILL ITS AUTHOR BE LIABLE FOR DAMAGES ARISING DIRECTLY OR INDIRECTLY FROM ANY USE OF THE INFORMATION CONTAINED IN THIS DOCUMENT.

Verilog is a registered trademark of Cadence Design Systems, Inc.

This document specifies a preprocessor, primarily intended to supersede the existing Verilog/SystemVerilog one with a Turing-complete language. It is however compatible with any language that does not make use of the backtick.

The main text uses the following conventions:

* Bold is used for chapters and sections title formatting, and to highlight important information and document-specific terms.
* Monospace font is used for syntax descriptions and source code examples.
* Bold-red monospace characters denote reserved keywords, operators, and punctuation marks as a required part of the syntax.

The formal syntax is described using Backus-Naur Form (BNF). The following conventions are used:

* A vertical bar that is not in bold-red separates alternative items.
* Square brackets that are not in bold-red enclose optional items. The item may appear only once.
* Braces that are not in bold-red enclose a repeated item. The item may appear zero or more times; the repetitions occur from left to right as with an equivalent left-recursive rule.

Table of Contents

[1 Delimitation 5](#_Toc63790757)

[2 Keywords 6](#_Toc63790758)

[3 Expressions 7](#_Toc63790759)

[4 Variables 8](#_Toc63790760)

[4.1 Variables definition 8](#_Toc63790761)

[4.2 Dynamic naming 8](#_Toc63790762)

[4.3 String interpolation 9](#_Toc63790763)

[5 Datatypes 10](#_Toc63790764)

[5.1 Integers 10](#_Toc63790765)

[5.2 Booleans 10](#_Toc63790766)

[5.3 Strings 10](#_Toc63790767)

[5.4 Arrays 10](#_Toc63790768)

[5.5 Dictionaries 11](#_Toc63790769)

[6 Statements 12](#_Toc63790770)

[6.1 Selection statement 12](#_Toc63790771)

[6.2 Iteration statement 12](#_Toc63790772)

[6.3 Expression statement 12](#_Toc63790773)

[6.4 Assertions 13](#_Toc63790774)

[7 Macros 14](#_Toc63790775)

[7.1 Macros definition 14](#_Toc63790776)

[7.2 Macros invocation 14](#_Toc63790777)

[8 Built-in Functions 15](#_Toc63790778)

[9 Files 16](#_Toc63790779)

[10 run 17](#_Toc63790780)

[A Example 18](#_Toc63790781)

# Delimitation

The preprocessor uses the backtick as a code delimiter, while the Verilog/SystemVerilog is seen as plain text. Except for evaluations, all preprocessor statements start at the backtick and end at the next, ending newline. Newlines appearing in between parentheses, brackets and curly braces, are not ending newlines.

A pass-through is given for Verilog/SystemVerilog directives by appending a backtick before theirs. The following character sequence is then seen as plain text, and reproduced in the output file. In that case, the extra backtick is removed.

**Example:**

``timescale 1ns/1ps

Here, the timescale directive is ignored by this preprocessor, and copied to the output file.

# Keywords

Keywords are predefined, non-escaped identifiers that are used to define the language constructs. All keywords are defined in lowercase only:

**assert**

**clog2**

**define**

**else**

**elseif**

**endfor**

**endif**

**endmacro**

**false**

**for**

**if**

**include**

**inside**

**let**

**log2**

**macro**

**max**

**min**

**size**

**true**

# Expressions

Expressions are sequences of operators and their operands that specify a computation. Operators, in the form of non-alphanumeric characters, always produce a result. Those symbols are listed below in the order of decreasing precedence; the highest precedence is listed first.

| Category | Symbol | Meaning | Associativity |
| --- | --- | --- | --- |
| Postfix | expression**[**expression**]**  expression**(**[expr {**,** expr}]**)** | Subscript  Macro call | Left to right |
| Prefix | **+**expression  **-**expression  **~**expression  **!**expression  **$**expression  **@**expression | Arithmetic identity  Arithmetic negation  Bitwise NOT  Logical NOT  String interpolation  Dynamic naming | Right to left |
|  | expression **\*\*** expression | Exponentiation | Right to left |
| Multiplicative | expression **\*** expression  expression **/** expression  expression **%** expression | Arithmetic product  Arithmetic quotient  Arithmetic remainder | Left to right |
| Additive | expression **+** expression  expression **–** expression | Arithmetic sum  Arithmetic difference | Left to right |
| Shift | expression **<<** expression  expression **>>** expression  expression **>>>** expression | Left shift  Logical right shift  Arithmetic right shift | Left to right |
| Relational | expression **inside** expression  expression **==** expression  expression **!=** expression  expression **<** expression  expression **<=** expression  expression **>** expression  expression **>=** expression | Set membership  Equal  Not equal  Less-than  Less-than or equal  Greater-than  Greater-than or equal | Not chainable |
| Bitwise | expression **&** expression | Bitwise AND | Left to right |
| expression **^** expression | Bitwise XOR | Left to right |
| expression **|** expression | Bitwise OR | Left to right |
| Logical | expression **&&** expression | Logical AND | Left to right |
| expression **||** expression | Logical OR | Left to right |
|  | expression **?** expr **:** expr | Ternary | Right to left |

# Variables

A variable is a storage paired with an identifier, which contains a quantum of information referred to as a value. An identifier is a sequence of one or more lowercase letters, digits and underscores, albeit the first character must not be a digit. A variable can be defined at most once. A definition binds an identifier to a type and assigns an initial value. Variables must be defined prior to their use in expressions.

## Variables definition

The syntax for defining a variable is:

**`define** expression **=** expression \n

with the left expression being declarable, that is, either an identifier or a dynamic name. Each variable is only valid in a continuous portion of the code called its scope. In that case, the scope extends from the point of definition to the end of the file in which they are defined.

Variables can be made local:

**`let** expression **=** expression \n

with the left expression being declarable, that is, either an identifier or a dynamic name. In that case, the scope extends from the point of definition to the end of the block in which it is defined. If there is a nested block with a definition that introduces identical identifier, the entire scope of the nested definition is excluded from the scope of the outer definition.

**Example:**

`let foo = 42

`define bar = "Hello"

In this example, variable “foo” is local while “bar” is made global.

## Dynamic naming

Variables can be dynamically named with the value stored in another variable. This is done with the at symbol:

**@**expression

The expression must be of string type. This operator is useful for deferred string interpolation in macros, where a name can be passed as argument.

**Example:**

`let variable\_name = "foo"

`let @variable\_name = 123

On the second line, a variable named “foo” containing the value “123” will be declared.

## String interpolation

String interpolation is a process substituting values of variables into placeholders in a string. String interpolation allows users to embed variable references directly in processed string literals:

**$**expression

The expression must be of string type. A string is returned, following the same resolution process as for the plain text. Note that no implicit interpolation is performed on string creation; they must be explicitly interpolated so that variables are replaced.

**Example:**

`let who = "world"

`let message = $"Hello `who!"

The variable “who” in the string will be substituted by its value “world”. The “message” variable will then contain “Hello world!”.

# Datatypes

The language supports the four base datatypes: integers, booleans, strings and arrays.

## Integers

Integers are literals made of a non-zero decimal digit followed by zero or more decimal digits. Presently, only decimal numbers are supported.

## Booleans

Boolean literals are the keywords **true** and **false**.

## Strings

Strings are literals made of any valid ASCII character, except the double-quote, and enclosed in between double-quotes. Special characters can be escaped with a backslash:

| **Escape sequence** | **Description** | **ASCII representation** |
| --- | --- | --- |
| \' | Single quote | 0x27 |
| \" | Double quote | 0x22 |
| \\ | Backslash | 0x5c |
| \a | Audible bell | 0x07 |
| \b | Backspace | 0x08 |
| \f | Form feed | 0x0c |
| \n | Line feed | 0x0a |
| \r | Carriage return | 0x0d |
| \t | Horizontal tab | 0x09 |
| \v | Vertical tab | 0x0b |

## Arrays

An array is an ordered set of items, which can be of different types. The array is identified with the name of its base variable. The syntax for defining an array is:

**[**[expression {**,** expression}]**]**

A continuous set of integers can be defined with the range operator:

expression **..** expression

where the left expression is the lower inclusive end of the range and the right expression is the upper inclusive end of the range. This operator can only appear inside an array definition as an expression.

**Example:**

`let array = [1..3, true, "Daffy Duck"]

Arrays elements are accessed through the subscript operator, with an integer index.

## Dictionaries

A dictionary is set of key-pair values, such that each possible key appears at most once in the collection. The dictionary is ordered against the key, not the insertion order like an array. The syntax for defining and array is:

**{**[expression **:** expression {**,** expression **:** expression}]**}**

with the expressions on the left of the colon being a string. Dictionaries elements are accessed through the subscript operator, with a string index.

**Example:**

`let money = {"Henry": 42, "Tom": 123}

`mut money["Tom"] += 10

# Statements

Statements are fragments of the code that are executed in sequence. The body of a program is a sequence of statements, though without explicit delimiters.

For further use, we define the following equivalence:

statements = {statement}

## Selection statement

Selection statements choose between one of several flows of control. The syntax is:

**`if (**expression**)** \n

statements

{**`elseif (**expression**)** \n

statements}

[**`else** \n

statements]

**`endif** \n

Statements are executed if the associated condition evaluates to a non-zero value. In that case, statements of the following conditions are skipped. If the value is zero, then the statements are skipped and the following, if present, are evaluated the same way. The “else” block has no condition and is always executed if not skipped. There can be zero or more “elseif” blocks followed by zero or one “else” block in each selection statement.

## Iteration statement

Iteration statements repeatedly execute the encompassed statements while successively assigning the values of the expression list to the variable defined by the identifier.

**`for (**expression **:** expression**)** \n

statements

**`endfor** \n

with the left expression being declarable, that is, either an identifier or a dynamic name; and the right expression being an array. In addition, the preprocessor records the current index in an “index” variable that is locally defined at each iteration.

## Expression statement

A single backtick followed by a keyword is a statement. A single backtick followed by an identifier is an expression statement. This statement evaluates the expression and substitutes it to the original text in the output file. While other statements end at the end-of-line, an expression statement ends at the first non-alphanumeric or underscore character, outside of any parentheses, brackets or curly braces group. The syntax is:

**`**expression

To allow for macro calls, if that character is an opening parenthesis, it is considered part of the expansion.

To prevent the preprocessor to consider an opening parenthesis, to expand in between an alphanumeric or underscore Verilog/SystemVerilog identifier, or to expand complex expressions, enclose the expression in between parenthesis as follow:

**`(**expression**)**

**Example:**

let lang = "verilog"

string my\_`(lang)\_variable;

In this example, the expanded text will be:

string my\_verilog\_variable;

## Assertions

Assertions are statements used to test assumptions made by programmer. They are mainly used to check logically impossible situations.

**`assert** **(**expression**)** \n

with the expression being of boolean type. If the argument expression compares equal to false, an error is generated. The preprocessor is not terminated, however, but the file is outputed.

# Macros

A macro associates to a set of values an output value with a sequence of statements. It can have zero or more input parameters but exactly one output.

## Macros definition

They are defined as follow:

**`macro** expression**(**[identifier {**,** identifier}]**)** \n

statements

**`endmacro** \n

with the left expression being declarable, that is, either an identifier or a dynamic name. The parameters of a macro definition are variables definition, with visibility restricted to the macro scope.

Macros are defined globally and have visibility on the caller scope. This removes the need for two-steps string interpolation, when some variables are defined in the caller’s scope, some in the callee’s one.

**Example:**

`macro say\_hello(name)

Hello `name

`endmacro

Called with “world” as parameter, this macro will print “Hello world” in the output file.

## Macros invocation

Macros are called as follow:

expression**(**[expression {**,** expression}]**)**

with the left expression being addressable, that is, either an identifier, a dynamic name, an array element, or a dictionary entry. When a macro is invoked, the parameters are initialized from the arguments and the statements in the macro body are executed.

Macros reproduce the plain text that they enclose, executing the statements inside, and performing the evaluations.

# Built-in Functions

The preprocessor features some built-in functions. The associated keywords are reserved.

| **Syntax** | **Description** |
| --- | --- |
| **log2(**expression**)** | Returns the ceiled logarithm in base 2 of the expression, which must be an integer. |
| **log2(**expression**)** | Returns the logarithm in base 2 of the expression, which must be an integer. |
| **min(**expression {**,** expression}**)** | Returns the minimum of the values given as arguments. |
| **max(**expression {**,** expression}**)** | Returns the maximum of the values given as arguments. |
| **size(**expression**)** | Returns the size of the array passed as argument. |

# Files

To be processed, files must have one of the two following extensions:

* “.src” for source files. They are compiled and a file is outputted for each one, in the same directory, and with the name but the extension removed. Preprocessor source files are usually Verilog/SystemVerilog source and header files.
* “.dat” for data files. They are compiled, but no file is outputted. Those are preprocessor exclusive files, such as libraries.

File inclusion can be used to insert the entire content of a source file in another file during generation. The result is as though the contents of the included source file appear in place of the statement. The syntax is as follow:

**`include** expression \n

with the expression being a string, representing the path from the working directory.

# run

To run the preprocessor, pass the files to process as follow:

./preprocess \*.src \*.dat

Presently, there is no option to set a destination folder. The output files are written in the same directory as their source file.

# Example

The following example is derived from Western Digital SweRV-EH1 core.

/// design/lsu/lsu\_bus\_intf.sv.src (original lines 305-365)

`include "libhdl.pph"

`for (i : [0..3])

`let bus = $"`((8\*i)+7):`(8\*i)"

`for (a : ["hi", "lo"])

`for (b : ["hi", "lo"])

`for (j : [3..5])

assign ld\_byte\_dc`(j)hit\_`(b)\_`(a)[`i] = ld\_addr\_dc`(j)hit\_`(b)\_`(a) &

ldst\_byteen\_`(b)\_dc`(j)[`i] & ldst\_byteen\_`(a)\_dc2[`i];

`endfor

`endfor

assign ld\_byte\_hit\_`(a)[`i] =

`or\_reduction("ld\_byte\_dc`(j)hit\_lo\_`(a)[`i]", "j", [3..5]) |

`or\_reduction("ld\_byte\_dc`(j)hit\_hi\_`(a)[`i]", "j", [3..5]) |

ld\_byte\_hit\_buf\_`(a)[`i];

`for (j : [3..5])

assign ld\_byte\_dc`(j)hit\_`(a)[`i] = ld\_byte\_dc`(j)hit\_lo\_`(a)[`i] |

ld\_byte\_dc`(j)hit\_hi\_`(a)[`i];

assign ld\_fwddata\_dc`(j)pipe\_`(a)[`bus] =

({8{ld\_byte\_dc`(j)hit\_lo\_`(a)[`i]}} & store\_data\_lo\_dc`(j)[`bus]) |

({8{ld\_byte\_dc`(j)hit\_hi\_`(a)[`i]}} & store\_data\_hi\_dc`(j)[`bus]);

`endfor

assign ld\_fwddata\_`(a)[`bus] = `decreasing\_prio("ld\_byte\_dc`(j)hit\_`(a)",

"ld\_fwddata\_dc`(j)pipe\_`(a)[`bus]", "ld\_fwddata\_buf\_`(a)", "j", [3..5]);

`endfor

`endfor

/// libhdl.dat

`macro or\_reduction(data, iter, range)

`for (@iter : range)

`if (index == 0)

( `$data

`else

| `$data

`endif

`endfor

)

`endmacro

`macro decreasing\_prio(select, data, data0, iter, range)

(

`for (@iter : range)

`$select ? `$data :

`endfor

`$data0 )

`endmacro

Note that the only portion of code reproduced here has been modified.