

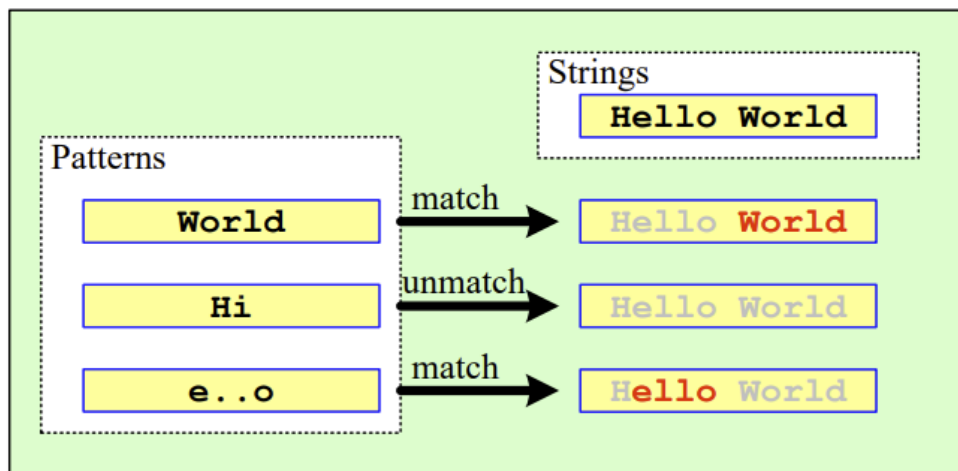
HOMEWORK 3 String Matching

Due date:

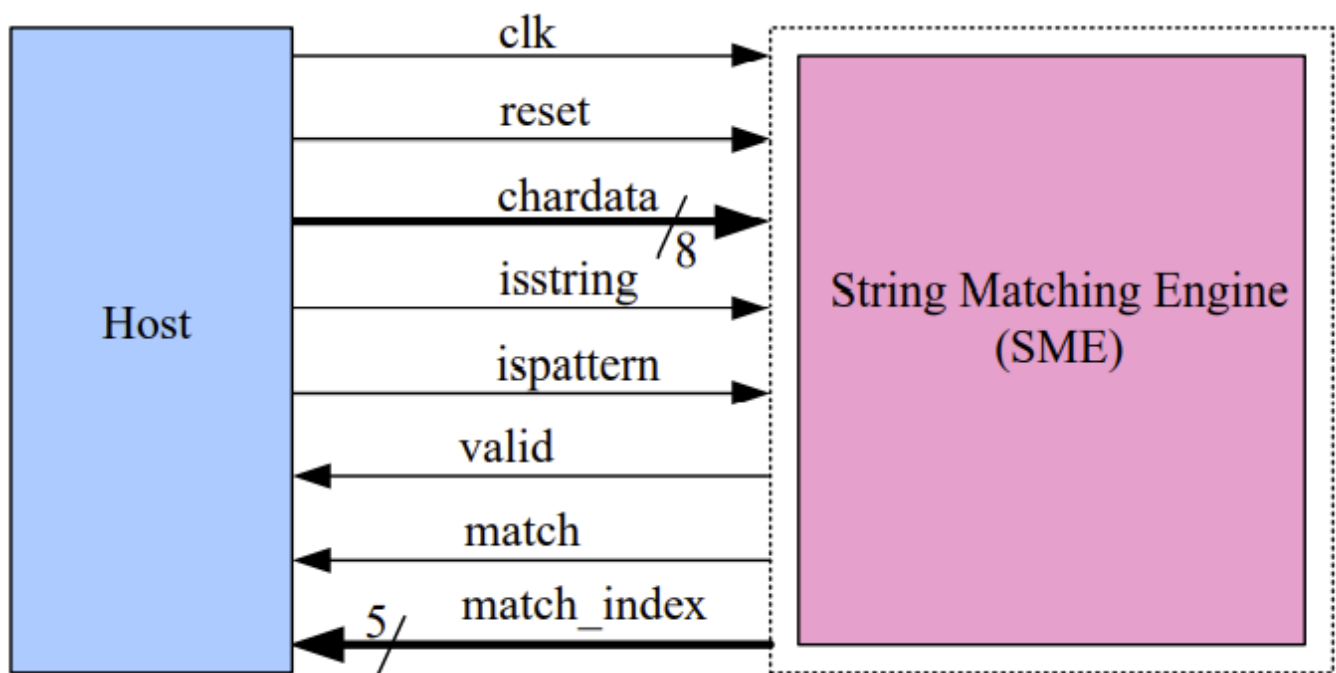
Overview

這次作業的主要目的是完成一 String Matching Engine (後文以 SME 表示)的電路設計，其功能為：將依序 提供數個字串(後文以 String 表示)及關鍵字樣本(後文以 Pattern 表示)，SME 電路需比對該 Pattern 是否包含在該 String 中，若有則回應比對成功(match)以及比對到的位置。

Homework Description



圖一、Pattern Match



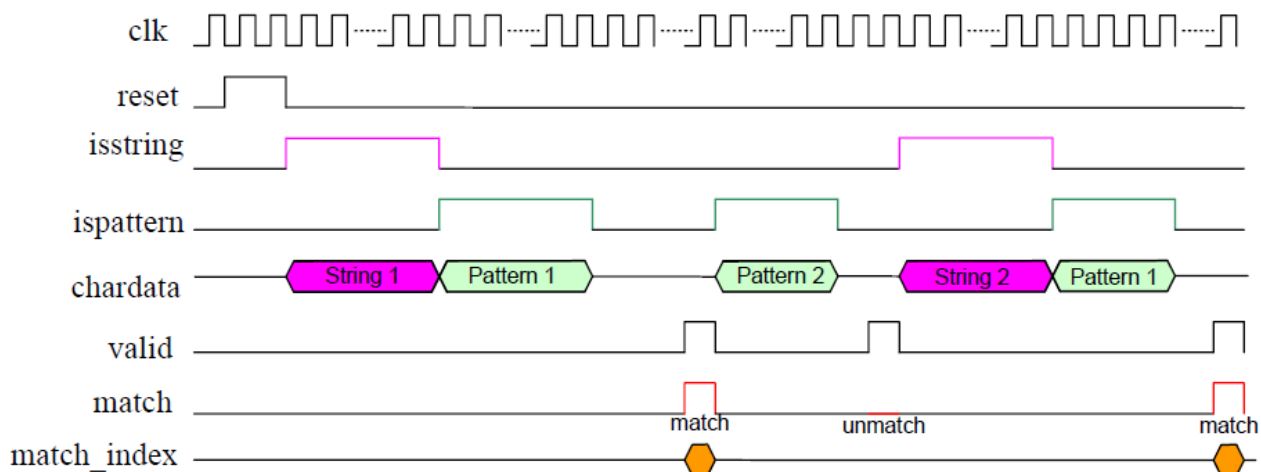
圖二、系統方塊圖

Signal Name	I/O	Width	Simple Description
clk	I	1	本系統為同步於時脈正緣之同步設計。
reset	I	1	高位準”非”同步(active high asynchronous)之系統重置信號。
chardata	I	8	輸入String 或Pattern，一次輸入一個ASCII 字元(8bit)
isstring	I	1	當isstring 為high 時，chardata 是屬於string 資料
ispattern	I	1	當ispattern 為high 時，chardata 是屬於pattern 資料
match	O	1	當string 和pattern 比對成功，輸出match 為high。
match_index	O	5	當string 和pattern 比對成功，輸出pattern 在string 中第一個比對成功的位置。
valid	O	1	有效的輸出訊號。當valid 為High，表示目前輸出的match、match_index 資料為有效的輸出，反之，當valid 為Low，表示match、match_index 為無效的輸出。

1. 系統描述

1.1 String 及 Pattern 輸入順序

String 及 Pattern 輸入順序採一個 String 搭配數個 Pattern 方式；String 及 Pattern 資料都從 chardata 輸入，一次輸入 1 個 ASCII 字元，當 isstring 為 high 時表示此時 chardata 輸入為 String 資料，當 ispattern 為 high 時表示此時 chardata 輸入為 Pattern 資料；String 和 Pattern 的長度非固定，但有上限，String 最長 32 字元，Pattern 最長 8 字元。每完成輸入一個 Pattern 後，測試端便會等待 SME 輸出比對結果，取得結果後立刻再輸入下一個 String 或 Pattern，請見圖三。



圖三、String 及 Pattern 輸入順序

1.2 SME 的輸出

當 ispattern 由 high 變為 low 時，表示該 Pattern 輸入完成，SME 可開始比對字串。比對完成後，請將 valid 訊號拉為 High，並在同一個 cycle 內，輸出 match 與 match_index 的比對結果。

若比對結果為不成功(match == 0)，則 match_index 不被參考，可為任意值。

1.3 Pattern 特殊符號說明

String 及 Pattern 每一字元以 ASCII 編碼，ASCII 編碼表請見附錄三 Pattern 中可能包含

以下四種特殊符號，特殊符號說明如下：

ASCII 字碼	符號	特殊符號說明
5E	^	比對word開頭
24	\$	比對word結尾
2E	.	比對任意單一字元

註：這三個特殊符號不會在 String 中出現。

上表 ” word” 指的是連續非空白的字元，比如說 String 內容是 “This is a pencil” 此字串內共含 4 個 “word”，分別是 “This”、“is”、“a”、“pencil”；

則以下 Pattern 皆可比對成功：

Pattern	match 字串	說明
^This	This is a pencil	^ 表示 T 開頭的 word
his\$	This is a pencil	\$ 表示 s 結尾的 word
^is\$	This is a pencil	^ 表示 I 開頭的 word, \$ 表示 s 結尾的 word
^a\$	This is a pencil	這 pattern 表示 word 裏只有一個 a 字元
^a pencil\$	This is a pencil	這 pattern 有兩個 word，最前面的 word 以 a 開頭，最後面的 word 以 l 結尾

且以下 Pattern 皆比對不成功：

Pattern	match 字串	說明
^his	This is a pencil	找不到 h 開頭的 word
pen\$	This is a pencil	找不到 n 結尾的 word
hi\$	This is a pencil	找不到 i 結尾的 word

註：為簡化狀況，“^” 只會出現在 Pattern 的最前面，“\$” 只會出現在 Pattern 最後面，且這兩個特殊符號不會和空白相連。

點號 “.” 可代表任意單一字元，以上面 String 例子，以下 Pattern 可比對成功：

Pattern	match 字串	說明
h.s	This is a pencil	. 代表 i 字元
p.n.il	This is a pencil	兩個 . 分別代表 e 和 c 字元
is.a.pe..il	This is a pencil	四個 . 分別代表 2 個空白和 nc 兩字元

註：一個 Pattern 可能有多個點號 “.” 特殊符號

1.4 比對成功位置 (match_index)說明

String 及 Pattern 資料都從 chardata 輸入，一次輸入 1 個 ASCII 字元，皆由第 0 字元開始依序輸入，如底下 String，是先輸入 T 字元，最後輸入 l 字元

String 內容： This is a pencil

index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
字元	T	h	i	s		i	s		a		p	e	n	c	i	l

當比對成功時，match_index 輸出 Pattern 在 String 中 match 的第一個位置，若比對結果為 unmatched，則 match_index 不被參考，可為任意值。請見底下例子：

編號	Pattern	是否 match	match 部位	match_index
1	pencil	match	This is a pencil	10
2	is	match	This is a pencil	2
3	^is	match	This is a pencil	5
4	h.s	match	This is a pencil	1
5	<u>pen</u>	match	This is a pencil	9
6	pen\$	unmatch	This is a pencil	--
7	^an	unmatch	This is a pencil	--

註：編號 2 Pattern 在 String 中有多種 match 可能，取 match_index 最小的結果。

註：編號 5 Pattern 最前方有一空白字元，因此 match_index 為第 9 字元。

2. 評分標準

本題不限制 clock 週期時間，可自行調整 clock 週期時間。

分數分為 2 部分，coding style 及 match score。Total score 公式為（100 分）：

$$\text{total score} = \text{coding style} * 10\% + \text{match score} * 90\%$$

coding style 佔總分的 10%，會檢查 Verilog 的 coding style，必須使用 2C1S coding style，

match score 佔總分的 90%，公式為：

- 若該 pattern 正確結果為 match，SME 輸出 match 且 match_index 正確，加 3 分
- 若該 pattern 正確結果為 match，SME 輸出 match 但 match_index 錯誤，加 1 分
- 若該 pattern 正確結果為 match，SME 輸出 unmatched，不計分
- 若該 pattern 正確結果為 unmatched，SME 輸出 unmatched，加 1 分
- 若該 pattern 正確結果為 unmatched，SME 輸出 match，不計分

General rules for deliverables

- You need to complete this homework **INDIVIDUALLY**. You can discuss the homework with other

students, but you need to do the homework by yourself. You should not **copy** anything from someone else, and you should not **distribute** your homework to someone else. If you violate any of these rules, you **will get NEGATIVE scores, or even fail this course directly**

- When submitting your homework, compress all files into a single **zip** file, and upload the compressed file to Moodle.

■ Please follow the file hierarchy shown in Figure 1.

F740XXXXX (your id) (folder)

src(folder) * Store your source code

report.docx (project report. The report template is already included. Follow the template to complete the report.)

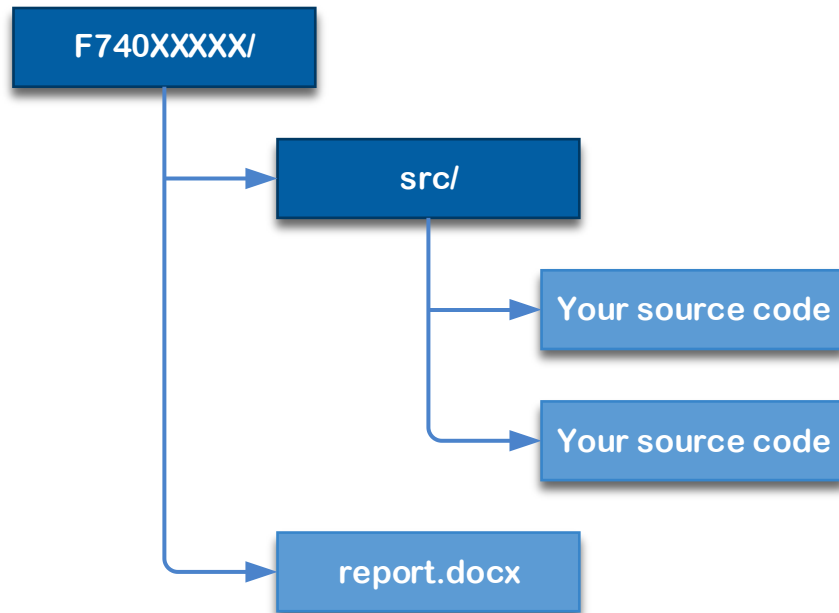


Figure 1. File hierarchy for homework submission

- **Important! DO NOT** submit your homework in the last minute. Late submission is not accepted.
- You should finish **all the requirements (shown below)** in this homework and Project report.
- **If your code can not be recompiled by TA successfully using modelsim, you will receive NO credit.**
- Verilog and SystemVerilog generators aren't allowed in this course.