Memory compiler tutorial – TSMC 40nm technology

2020/07/24 Shih Wei Hsieh

Outline

- Memory Compiler Introduction
- Memory Compiler Interface
 - Parameters
 - Files Generation
 - Using the Files
 - Memory Ports
- Lab #1 Dual-Port SRAM
- Lab #2 Asynchronous FIFO

Memory Compiler Introduction

- Memory compiler is the SRAM/ Register File/ ROM generator for ASIC design. The generated files include:
 - User guide
 - Verilog model (for simulation)
 - Timing information (for synthesis and APR)
 - Physical layout (for APR)
- We will focus on the Memory compiler for TSMC 40nm technology
 - Settings should be similar for other technology nodes
- Location:
 - 90nm: /cad/cell_library/CBDK_TSMC90G_Arm/CIC/Memory/
 - 40nm: /cad/cell_library/CBDK_TSMC40_Arm_f2.0/CIC/Memory/

Memory Compiler Introduction

• Folders:

Directory name	Generate type
rf_2p_hse_rvt_hvt_rvt/	High-speed two-port register file
rf_sp_hde_rvt_hvt_rvt/	High-density single-port register file
rf_sp_hsd_rvt_rvt_hvt/	High-speed single-port register file
rom_via_rvt_hvt_rvt/	ROM
sram_dp_hde_rvt_hvt_rvt	High-density dual-port SRAM
sram_sp_hde_rvt_hvt_rvt/	High-density single-port SRAM
sram_sp_hsc_rvt_hvt_rvt/	High-speed single-port SRAM

• Single-port: 1 read or 1 write per clock

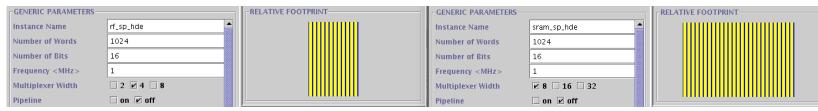
• Two-port: 1 read or 1 write or 1 read+ 1 write per clock

• Dual-port: 2 reads or 2 writes or 1 read+ 1 write per clock

Memory Compiler Introduction

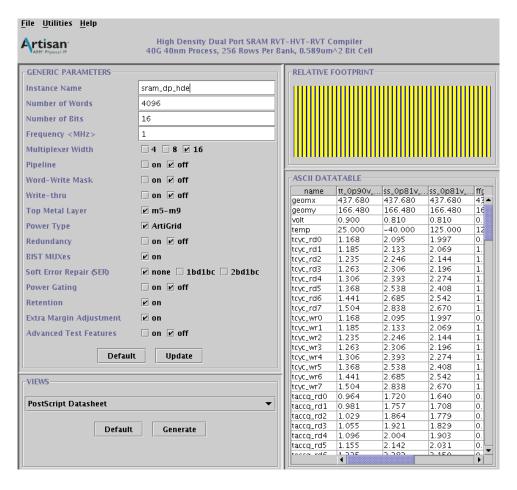
- Difference between SRAM and RF?
 - RF is usually used for smaller memory (lower address line bits)
- E.g. Single-port SRAM vs single-port RF
 - SRAM number of words range: [256, 16384]
 - RF number of words range: [16, 2048]
- For the number of words in the overlapped range (e.g. 1024 words)

- Layout comparison: RF SRAM

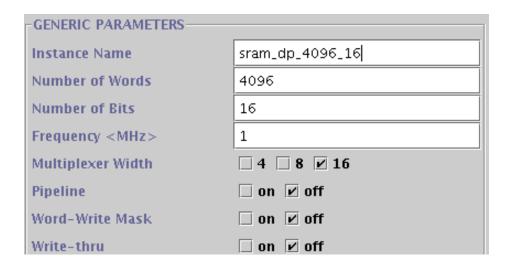


- Area comparison:
 - RF mux width (4, 8): $(10291, 10165) \mu m^2$
 - SRAM mux width (8, 16, 32): (13346, 16651, 26693) μm^2

- Interface (take dual-port SRAM for example):
 - /cad/cell_library/CBDK_TSMC40_Arm_f2.0/CIC/Memory/sram_dp_hde_rvt_h vt_rvt/r5p0/bin/sram_dp_hde_rvt_hvt_rvt



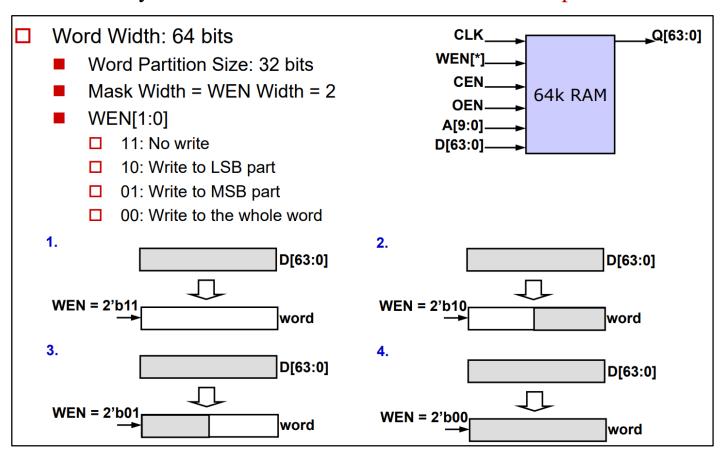
- Usually we only modify:
 - Name
 - Number of words
 - Number of bits
 - Multiplexer Width
 - (Word-write mask)



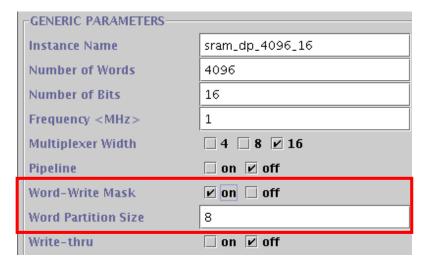
- Multiplexer width:
 - You should not read and write in the same clock where the address difference < multiplexer width to avoid row contention
 - Otherwise, the read or write operation might fail

```
row contention: in sram tb.sram_unit.genBlock1[1].sbj_memory at 194000 contention: read B succeeds in sram_tb.sram_unit.genBlock1[1].sbj_memory at 194000 contention: write A fails in sram_tb.sram_unit.genBlock1[1].sbj_memory at 194000
```

- Word-write mask:
 - Useful when you don't want to write the whole word but part of the word



- Word-write mask:
 - Useful when you don't want to write the whole word but part of the word
 - Word partition option will appear after you check the word-write box

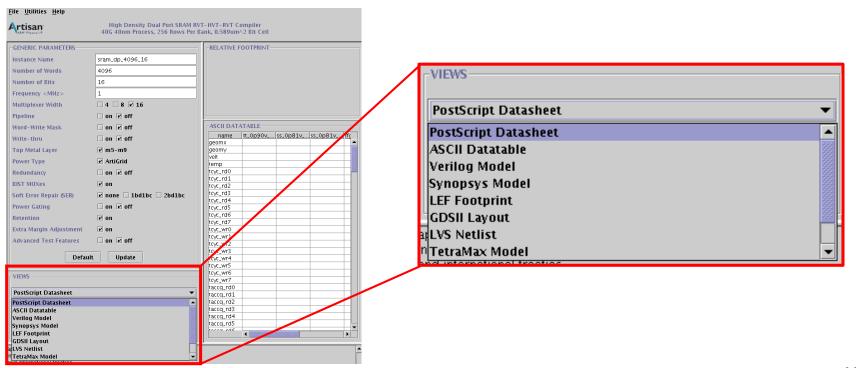


- Without word-write mask, you have to do the following steps to achieve the same result:
 - Read the content first \rightarrow change part of the data \rightarrow write the whole word back

- Pipeline:
 - A flip-flop is placed between the output of the memory and QA
- Write-through:
 - The data written into the memory is propagated through to the output port
- If you are interested in other options, you can find the detail explanation in the user guide document located at */Memory/*/*/doc/*_userguide.pdf
 - E.g.

/cad/cell_library/CBDK_TSMC40_Arm_f2.0/CIC/Memory/sram_dp_hde_rvt_hv t_rvt/r5p0/doc/sram_dp_hde_rvt_hvt_rvt_userguide.pdf

- You will only need:
 - Verilog model (.v): for neverilog simulation
 - Synopsys model (.lib): timing information for synthesis and APR tool
 - LEF footprint (.lef/.clf): physical layout for the APR tool
 - PostScript DataSheet (.ps) (optional): user guide

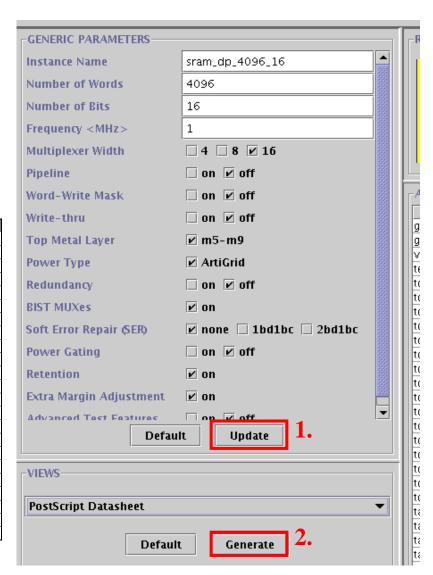


- PostScript DataSheet (.ps)
 - The generated files are *ps* files
 - Convert to PDF: ps2pdf *.ps
 - You can find the pin description in the document

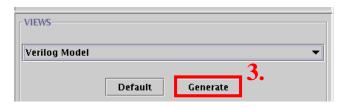
Pin Description

Pin	Description
AA[5:0], AB[5:0]	Port A & B Addresses (AA[0],AB[0] = LSB)
DA[127:0], DB[127:0]	Port A & B Data Inputs (DA[0],DB[0] = LSB)
CLKA, CLKB	Port A & B Clocks
CENA, CENB	Port A & B Chip Enables (active low)
WENA,WENB	Port A & B Write Enables (active low)
QA[127:0], QB[127:0]	Port A & B Data Outputs (QA[0],QB[0] = LSB)
EMAA[2:0], EMAB[2:0]	Port A & B Margin Adjustment (EMAA[0],EMAB[0] = LSB)
EMASA, EMASB	Port A & B Sense Amp Extra Margin Adjustment (EMASA,EMASB)
EMAWA[1:0], EMAWB[1:0] = LSB	Port A & B Write Extra Margin Adjustment (EMAWA[0],EMAWB[0])
TENA, TENB	Port A & B Test Mode Enables (active low)
TAA[5:0], TAB[5:0]	Port A & B Address Test Inputs (TAA[0],TAB[0] = LSB)
AYA[5:0], AYB[5:0]	Port A & B Address Mux Outputs (AYA[0],AYB[0] = LSB)
TDA[127:0], TDB[127:0]	Port A & B Data Test Inputs (TDA[0],TDB[0] = LSB)
DYA[127:0], DYB[127:0]	Port A & B Data Mux Outputs (DYA[0],DYB[0] = LSB)
TCENA, TCENB	Port A & B Chip Enable Test Inputs
CENYA, CENYB	Port A & B Chip Enable Mux Outputs
TWENA,TWENB	Port A & B Write Enable Test Inputs (active low)
WENYA,WENYB	Port A & B Write Enable Mux Outputs
BENA, BENB	Port A & B Bypass Mode Enables (active low)
TQA[127:0], TQB[127:0]	Port A & B Test mux QA,QB Inputs (TQA[0],TQB[0] = LSB)
COLLDISN	Allow the user to disable the internal collision detection circuitry(active low)
RET1N	Retention Input (active low)
STOVA, STOVB	Self timing override inputs

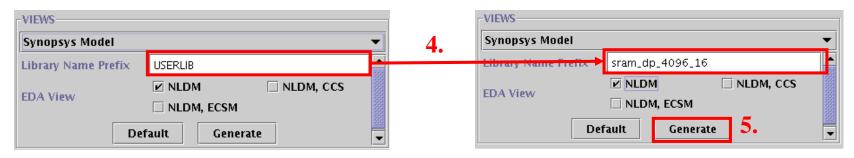
Pin description for the DP-SRAM



Verilog model



- Synopsis model
 - The default library name is "USERLIB"
 - It is recommended that you change the name to your instance name



- Use lc_shell to convert .lib to .db (necessary)
- LEF Footprint



- Use lc_shell to convert .lib to .db (necessary)
 - Step.1 Open library compiler with lc_shell
 - Step.2 Read library (read_lib *.lib)

```
>> read_lib sram_dp_64x128_nldm_tt_0p90v_0p90v_25c_syn.lib
```

Library file

```
Technology library 'sram_dp_64x128_nldm_tt_0p90v_0p90v_25c' read successfully 1 lc_shell> Library name (will be "USERLIB_..." if you didn't change the name)
```

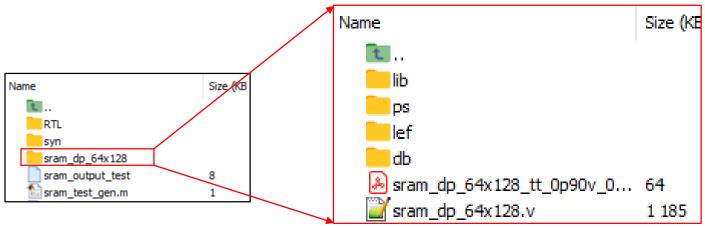
Step.3 Convert to db (write_lib * -o *.db)

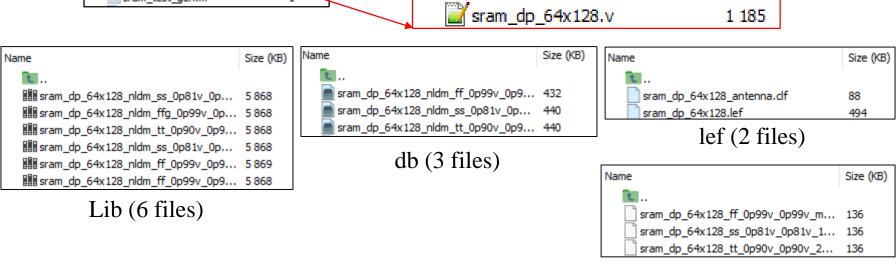
```
>> write_lib sram_dp_64x128_nldm_tt_0p90v_0p90v_25c
-o sram_dp_64x128_nldm_tt_0p90v_0p90v_25c_syn.db
```

db file

- We will convert three different corners
 - Slow: ss + 125c
 - Typical: tt + 25c
 - Fast: ff + m40c

- Organizing your files
 - Move files to their respective folders (lib/ ps/ lef/ db)





Memory Compiler – Using the files

- Verilog model: post rtl/syn/apr simulation
 - ncverilog ... sram_dp_64x128.v
- DB files (Synthesis):
 - Add path to .*synopsys_dc.setup*

Memory Compiler – Using the files

- Library files (APR):
 - Add path to the .view file

```
create library set -name lib max \
             -timing {/cad/CBDK/CBDK TN40G Arm/CBDK TSMC40 core Arm v2.0/CIC/SynopsysDC/lib/sc9 base rvt/sc9 cln40g base rvt ss typical max 0p81v 125c.lib \
             /cad/CBDK/CBDK TN40G Arm/CBDK TSMC40 core Arm v2.0/CIC/SynopsysDC/lib/sc9 base lvt/sc9 cln40g base lvt ss typical max 0p81v 125c.lib \
             /cad/CBDK/CBDK TN40G Arm/CBDK TSMC40 core Arm v2.0/CIC/SynopsysDC/lib/sc9 base hvt/sc9 cln40g base hvt ss typical max 0p81v 125c.lib \
             ../blackScholes/sram dp 1024 120/lib/sram dp 1024 120 nldm ss 0p81v 0p81v 125c syn.lib
             ../blackScholes/sram dp 2048 120/lib/sram dp 2048 120 nldm ss 0p81v 0p81v 125c syn.lib \
             ../blackScholes/sram sp 256 120/lib/sram sp 256 120 nldm ss 0p81v 0p81v 125c syn.lib \
             ../blackScholes/sram sp 4096 120/lib/sram sp 4096 120 nldm ss 0p81v 0p81v 125c syn.lib \
             ../blackScholes/rf 2p 16 2407lib/rf 2p 16 240 nldm ss 0p81v 0p81v 125c syn.lib \
             -si {/cad/CBDK/CBDK_TN40G_Arm/CBDK_TSMC40_core_Arm_v2.0/CIC/S0CE/celtic/sc9_base_rvt/sc9_cln40g_base_rvt_ss_typical_max_0p81v_125c.cdB_\
             /cad/CBDK/CBDK TN40G Arm/CBDK TSMC40 core Arm v2.07CIC/SOCE/celtic/sc9 base lvt/sc9_cln40g_base_lvt_ss_typical_max_0p81v_125c.cdB \
             /cad/CBDK/CBDK_TN40G_Arm/CBDK_TSMC40_core_Arm_v2.0/CIC/SOCE/celtic/sc9_base_hvt/sc9_cln40g_base_hvt_ss_typical_max_0p81v_125c.cdB_\
create library set -name lib min \
             -timing {/cad/CBDK/CBDK TN40G Arm/CBDK TSMC40 core Arm v2.0/CIC/SynopsysDC/lib/sc9 base rvt/sc9 cln40g base rvt ff typical min 0p99v m40c.lib \
             /cad/CBDK/CBDK TN40G Arm/CBDK TSMC40 core Arm v2.0/CIC/SynopsysDC/lib/sc9 base lvt/sc9 cln40g base lvt ff typical min 0p99v m40c.lib \
            /cad/CBDK/CBDK TN40G Arm/CBDK TSMC40 core Arm v2.0/CIC/SynopsysDC/lib/sc9 base hvt/sc9 cln40g base hvt ff typical min 0p99v m40c.lib \
             ../blackScholes/sram dp 1024 120/lib/sram dp 1024 120 nldm ff 0p99v 0p99v m40c syn.lib \
             ../blackScholes/sram dp 2048 120/lib/sram dp 2048 120 nldm ff 0p99v 0p99v m40c syn.lib \
             ../blackScholes/sram sp 256 120/lib/sram sp 256 120 nldm ff 0p99v 0p99v m40c syn.lib \
             ../blackScholes/sram sp 4096 120/lib/sram sp 4096 120 nldm ff 0p99v 0p99v m40c syn.lib \
             ../blackScholes/rf 2p 16 240/lib/rf 2p 16 240 nldm ff 0p99v 0p99v m40c syn.lib \
             -si {/cad/CBDK/CBDK TN40G Arm/CBDK TSMC40 core Arm v2.0/CIC/SOCE/celtic/sc9 base rvt/sc9 cln40g base rvt ff typical min 0p99v m40c.cdB ∖
             /cad/CBDK/CBDK TN40G Arm/CBDK TSMC40 core Arm v2.0/CIC/SOCE/celtic/sc9 base lvt/sc9 cln40g base lvt ff typical min 0p99v m40c.cdB \
             /cad/CBDK/CBDK_TN40G_Arm/CBDK_TSMC40_core_Arm_v2.0/CIC/SOCE/celtic/sc9_base_hvt/sc9_cln40g_base_hvt_ff_typical_min_0p99v_m40c.cdB \
```

Memory Compiler – Using the files

- Lef files (APR):
 - Add path to the .globals file

- APR for 40nm technology requires access to the EDA Cloud
 - We will not practice APR today

Memory Compiler – Ports

• Single-port SRAM:

Port name	Description
CLK	Clock
CEN	Active-low chip enable (0 to enable)
WEN	Active-low write enable (1 for read, 0 for write)
A	Address
D	Data input
Q	Data output
Ohers	Connect to 0 or 1

```
sram sp 4096 120 UX reg mem4 (
    .CLK (CLK),
    .CEN( !(UX w en4||UX r en4) ),
    .WEN(!UX w en4),
    .A(UX addr4),
    .D(UX in4[((idx+1)*120-1)-:120]),
    // output
    //.QA(),
    .Q(UX out4[((idx+1)*120-1)-:120]),
    .EMA (3'b0),
    .EMAS (1'b0),
    .EMAW (2 b0),
    .BEN (1'b1),
    .STOV (1'b0),
    .TEN (1'b1),
    .TCEN (1'b1),
    .TWEN (1'b1),
    .TA(12'b0),
    .TD(120'b0),
    .TQ(120'b0),
    .RET1N(1'b1)
```

Example

Memory Compiler – Ports

• Dual-port SRAM:

Port name	Description
CLKA/ CLKB	Clock
CENA/ CENB	Active-low chip enable (0 to enable)
WENA/ WENB	Active-low write enable (1 for read, 0 for write)
AA/ AB	Address
DA/ DB	Data input
QA/ QB	Data output
Ohers	Connect to 0 or 1

```
sram dp 64x128 sbj memory ( // can store at m
    .CLKA (CLK),
    .CLKB (CLK),
    .CENA (!sram WEN),
    .CENB(!sram REN ),
    .WENA (1'b0),
    .WENB (1'b1),
    .AA (sram waddr),
    .AB (sram raddr),
    .DA(sram in[(128*(gen idx+1)-1)-:128]),
    .DB (128'b0),
    // output
    //.QA(),
    .QB(sram out[(128*(gen idx+1)-1)-:128]),
    .EMAA (3'b0),
    .EMAB (3'b0),
    .EMASA (1 b0),
    .EMASB (1'b0),
    .EMAWA (2 b0),
    .EMAWB (2 b0),
    .BENA (1'b1),
    .BENB (1'b1),
    .STOVA (1'b0),
    .STOVB (1'b0),
    .TENA (1'b1),
    .TENB (1'b1),
    .RET1N (1'b1)
```

Example (always 1 read 1 write)

Memory Compiler – Ports

• 2-port RF:

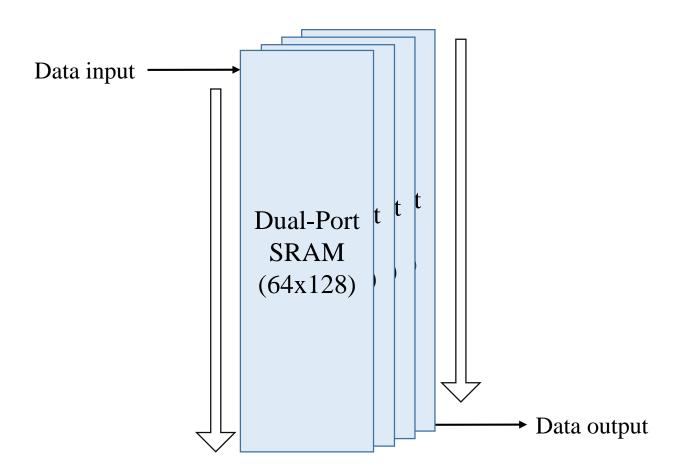
Port name	Description
CLKA/ CLKB	Clock
CENA/ CENB	Active-low chip enable (0 to enable)
AA/ AB	Address
DB	Data input
QA	Data output
Ohers	Connect to 0 or 1

```
rf 2p 16 240 last UX mem (
    .CLKA (CLK),
    .CLKB (CLK),
    .CENA( !last UX r en ),
    .CENB(!last UX w en ),
    .AA(last UX out addr),
    .AB(last UX in addr),
    .DB(last UX in[idx]),
    .QA(last UX out[idx]),
    .EMAA (3'b0),
    .EMAB (3'b0),
    .EMASA (1 b0),
    .EMAWB (2 b0),
    .BENA (1'b1),
    .STOVA (1'b0),
    .STOVB (1'b0),
    .TENA (1'b1),
    .TENB (1'b1),
    .RET1N(1'b1),
    .TCENA (1'b1),
    .TAA (4'd0),
    .TQA (240'd0),
    .TCENB (1'b1),
    .TAB (4'd0),
    .TDB (240'd0),
    .COLLDISN (1'b1)
```

Example (A: read; B: write)

Lab #1 – Dual-Port SRAM

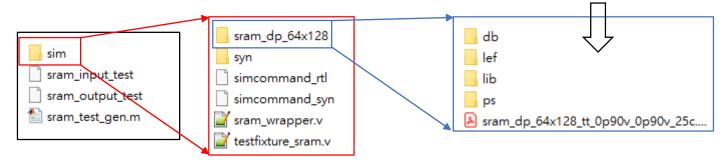
- Function:
 - Input \rightarrow Store in SRAM \rightarrow Read from SRAM \rightarrow Output



Lab #1 – Dual-Port SRAM

- We have completed the RTL files and synthesis script
- Your task:
 - Use the memory compiler to generate the necessary files and put them in respective folders

Fill the blank folders with the .db/.lib/.lef/.ps/.v files



- Spec:
 - Instance name: sram_dp_64x128
 - Number of words: 64
 - Number of bits: 128
 - Multiplexer width: 4

Lab #1 – Dual-Port SRAM

• Steps:

- Generate all the files
- 2. RTL simulation:
- 3. Synthesis (syn/):
- 4. Post-Synthesis simulation:



Pass simulation

	bash simco	nman	d_syn	ı	
	Group (max_delay/setup)	Cost	Weight	Weighted Cost	
	CLK default	0.00 0.00	1.00 1.00	0.00 0.00	
	max_delay/setup			0.00	
	Group (critical_range)	Total Neg Slack		s Cost	
	CLK default	0.00 0.00	0 0	0.00 0.00	
	critical_range			0.00	
	Group (min_delay/hold)	Cost	Weight	Weighted Cost	
	CLK default	0.00 0.00	1.00 1.00	0.00 0.00	
	min_delay/hold			0.00	
	Constraint			Cost	
	multiport_net min_capacitance max_transition max_fanout max_delay/setup sequential_clock_pulse_wi critical_range min_delay/hold min_period	dth		0.00 (MET) 0.00 (MET) 0.00 (MET) 0.00 (MET) 0.00 (MET) 0.00 (MET) 0.00 (MET) 0.00 (MET) 0.00 (MET)	
1	max_area		2	262186.38	

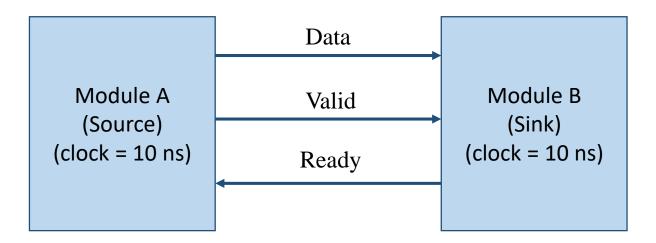
bash simcommand_rtl

dc_shell -f all_syn.tcl

Synthesis result

(VIOLATED)

- Data transfer between hardware modules
- The most frequently used handshake mechanism in digital design is the valid-ready protocol



- Source asserts valid when it has data to transfer, and puts data on the line
- Sinks asserts ready when it is ready to receive, and stores the data
- Source acknowledges the ready signal and sends the next data
- What if the source generates the data faster than the sink can consume them?
 - E.g. Source generates data in a burst while sink receives them slow but steadily

- Data transfer between hardware modules
- A synchronous FIFO can be inserted that act as a buffer when the data transfer is not at the same rate



- Synchronous: the same clock is used for both reading and writing
- FIFO depth: the size of the buffer inside the FIFO
 - Calculating FIFO depth: http://www.asic-world.com/tidbits/fifo_depth.html
- What if source and sink operate at different frequency?

- Data transfer between hardware modules
- An asynchronous FIFO uses different clock for reading and writing



• The design tips of asynchronous FIFO can be found in paper:

Cummings, Clifford E. "Simulation and synthesis techniques for asynchronous FIFO design." *SNUG 2002 (Synopsys Users Group Conference, San Jose, CA, 2002) User Papers.* 2002.

 We will be using the implementation from https://github.com/dpretet/async_fifo for today's lab

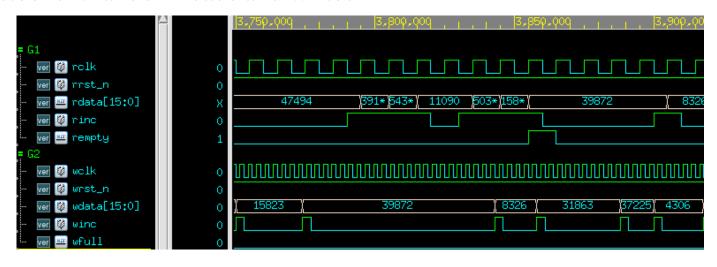
• Asynchronous FIFO ports (2⁸ words, 16 bits data):

```
module async fifo
#(parameter DSIZE = 16, parameter ASIZE = 8,
                                            parameter FALLTHROUGH = "FALSE")
                                             Read data will appear in the next clock
                            wclk,
    input wire
    input wire
                            wrst n,
    input wire
                            winc,
    input wire [DSIZE-1:0] wdata,
    output wire
                            wfull.
    output wire
                            awfull,
                            rclk,
    input wire
    input wire
                            rrst n,
    input wire
                            rinc,
    output wire [DSIZE-1:0] rdata,
    output wire
                            rempty,
   output wire
                            arempty
);
```

Port name	Description	
wclk	Clock	
wrst_n	Active low reset	
winc	Write enable	
wdata	Write data	
wfull	Indicates FIFO is full	
awfull	Almost full	

Port name	Description	
rclk	Clock	
rrst_n	Active low reset	
rinc	Read enable	
rdata	Read data	
rempty	Indicates FIFO is empty	
arempty	Almost empty	

• Testbench: random reads and writes



- Default: balanced number of reads and writes
- +define+PAT2: increase number of writes (test full)



- +define+PAT3: decrease number of writes (test empty)



- Data memory inside the FIFO (fifo_2mem.v):
 - The provided file uses registers to compose the buffer
 - For FPGA synthesis, this will be synthesized as block RAM by inference
 - For ASIC design, however, this will be synthesized as registers (large area)
 - We can replace the memory with a 2 port RF to reduce the area

```
reg [DATASIZE-1:0] mem [0:DEPTH-1];
  always @(posedge wclk) begin
      if (wclken && !wfull)
          mem[waddr] <= wdata;</pre>
  end
generate
  if (FALLTHROUGH == "TRUE")
    begin : fallthrough
      always @*
        rdata = mem[raddr];
    end
  else
    begin : registered read
      always @(posedge rclk) begin
        if (rclken)
          rdata <= mem[raddr];
      end
    end
endgenerate
```

• You should be able to successfully run simulation and synthesis before any modification:

- RTL simulation: bash simcommand_rtl

- Synthesis (syn/): dc_shell -f all_syn.tcl

- Post-Synthesis simulation: bash simcommand_syn

• The area of FIFO after synthesis is 26856

```
      Combinational area:
      10653.249416

      Buf/Inv area:
      425.023203

      Noncombinational area:
      16202.819267

      Macro/Black Box area:
      0.000000

      Net Interconnect area:
      0.000000

      Total cell area:
      26856.068682

      Total area:
      26856.068682
```

- In this lab, You task is to replace the FIFO memory (fifo_2mem.v) with a 2-port register file to reduce the total cell area to < 10000
- Step-by-step hints:
 - Find the proper parameters (# of words, # of bits, ...) for the 2-port RF
 - Generate all the files (.db, .lib, .v)
 - Replace the memory in *fifo_2mem.v* with the RF you generates
 - Modify *simcommand_rtl* and run RTL simulation
 - Add the path of the db files to .synopsys_dc.setup
 - Run synthesis (record total cell area)
 - Modify *simcommand_syn* and run post-synthesis simulation
 - Done

```
      Combinational area:
      230.428797

      Buf/Inv area:
      23.814000

      Noncombinational area:
      344.962783

      Macro/Black Box area:
      9221.239258

      Net Interconnect area:
      0.000000

      Total cell area:
      9796.630839

      Total area:
      9796.630839
```

Reference result