

1-D Time-Domain Convolution using FPGA

Priyanka Abhijit Tamhankar
UFID - 40893970
*Department of Electrical
and Computer Engineering*
University of Florida
Gainesville FL, USA
ptamhankar@ufl.edu

Meghana Koduru
UFID - 64766702
*Department of Electrical
and Computer Engineering*
University of Florida
Gainesville FL, USA
meghanakoduru@ufl.edu

Abstract – In this project a FPGA is programmed to perform 1-D time domain convolution using VHDL. It consists of two major parts to be completed which were the DMA interface between memory(DRAM) and the user_app and the design of the signal and kernel buffers to perform convolution.

I. INTRODUCTION:

Time domain convolution is a very common operation in image and digital signal processing applications. Convolution function takes in two signals, one of which is the input signal and the other is the filter function or kernel function. Convolution represents how the input varies as it's passed through or slide through the filter function. As this function has a series of repetitive multiplications and additions, it has a scope for huge amount of parallelism and hence, can be implemented on FPGA to achieve good performance.

II. DESIGN AND IMPLEMENTATION:

A. DMA INTERFACE :

The DMA controller serves the purpose of passing blocks of data between input/output device ports without minimal intervention of processor. In this project, A DMA controller is designed to read data from the DRAM (RAM0) and pass it to the USER_APP interface for computation. It has to serve various purposes like generating addresses to indicate which memory address the data has to be read from and synchronization of signals as the data has to be passed across different clock domains. The DRAM is running at a frequency of 133Mhz and the USER_APP is running at a frequency of 100 Mhz.

1. FIFO BUFFER

The data read from the DRAM has to be passed through the DMA entity and it's read by USER_APP for computation. The FIFO acts as a buffer to synchronize data that is read from and memory and written into the pipeline at different clock frequencies. It serves the purpose of a buffer and also avoids metastability to occur. FIFO is used here as this is a case where high throughput is a requirement. The FIFO is designed to read a 32 bit word every time there is a valid address generated from the address generator and the read enable of the DRAM is enabled, and output a 16 bit word to USER_APP every time read enable is asserted from USER_APP.

2. ADDRESS GENERATOR

When go signal of DMA interface is asserted, it triggers the address generator to produce addresses starting from the input START_ADDRESS until SIZE addresses are generated. The address generation needs to be stalled whenever the capacity of FIFO is full or whenever The DRAM_READY signal is zero which signals that there is no valid data in DRAM to be read. One important thing that needs to be considered is that the size from the USER_APP corresponds to 16 bit words where as the one address in DRAM corresponds to a 32 bit word. Hence only size/2 addresses need to be generated for reading 'size' number of 16-bit data words.

3. HANDSHAKE SYNCHRONIZER

The signal that triggers the address generator to start generation of addresses is in a different clock domain than the USER_APP, hence it should be synchronized or else metastability can occur. A handshake synchronizer is used between the address generator and the USER_APP interface of the DMA. The size and start address values are stored in registers and passed to the address generator inputs only when the synchronized go triggers address generation.

4. COUNTER

The counter indicates that size data bits have been read from FIFO and DMA operation is completed for the particular case. The counter will count up to size every time valid data is being read from FIFO and the FIFO is not empty. After the count is incremented till size, done is asserted indicating that all the expected data has being read from FIFO.

5. REGISTERS, AND and NOT gates

The register entities are instantiated to store the values of size and start addresses until the go signal from the USER_APP interfaced in the DRAM clock domain. The AND and NOT gates are used to add additional logic for the flush signal of the DRAM and empty signal of FIFO

B. USERAPP

It consists of three components:

1. convolution pipeline
2. signal buffer
3. kernel buffer

1. CONVOLUTION PIPELINE:

Convolution pipeline is a multiplier-adder tree which takes two arrays of inputs. One input to the pipeline is the output from the signal buffer whereas the other is the output from the kernel buffer. The output of the pipeline is obtained by sum of products of multiplying kernel and signal elements. It consists of 128 multipliers and 127 adders in the adder tree. Inputs from kernel buffer stay the same throughout the process whereas from signal buffer, they shift by one as it acts as a sliding window.

We made sure to enable the pipeline only when RAM1 is ready.

MultAddTree_En \leftarrow ram1_wr_ready

Convolution code is obtained from the provided code in which we made a few modifications:

a. DELAY LOGIC

The output of the datapath goes to the RAM1 entity and hence there should be a way to know when the output of the pipeline is valid. For this reason, valid_in and valid_out are added. valid_in is asserted whenever signal buffer rd_en is '1'. When the data is valid and RAM1_wr is ready, then the output will be written to RAM1. Valid_out when asserted informs RAM1 that the data being sent is valid.

MultAddTree_Valin_In \leftarrow Signalbuffer_rd_en

ram1_wr_valid \leftarrow Datapath_Valid_Out and ram1_wr_ready

Furthermore, Valid_in is delayed by cycles equal to the depth or latency of the datapath.

Latency = 9 cycles (for multipliers + adder tree ----- gets reduced by half every cycle)

b. CLIPPING LOGIC

The output of the datapath is greater than 16 bits but we need to send 16 bit output to RAM1. For this purpose, clipping logic is created to clip the output whenever it is greater than 16 bits. The output of this logic is then connected to the DRAM1 interface. The logic is such that if the overflow bits are 0's then, the output of clipping logic will be the first 16 bits of datapath otherwise the output is all 1's.

The following screenshot shows the clipping logic:

→ As indicated by the yellow marker, the output from the datapath is 000000003c. This output is later later clipped to 003c. (ram1_wr_data).

→ As seen below, we also made sure that datapath enable is asserted only when ram1_wr_ready is asserted.

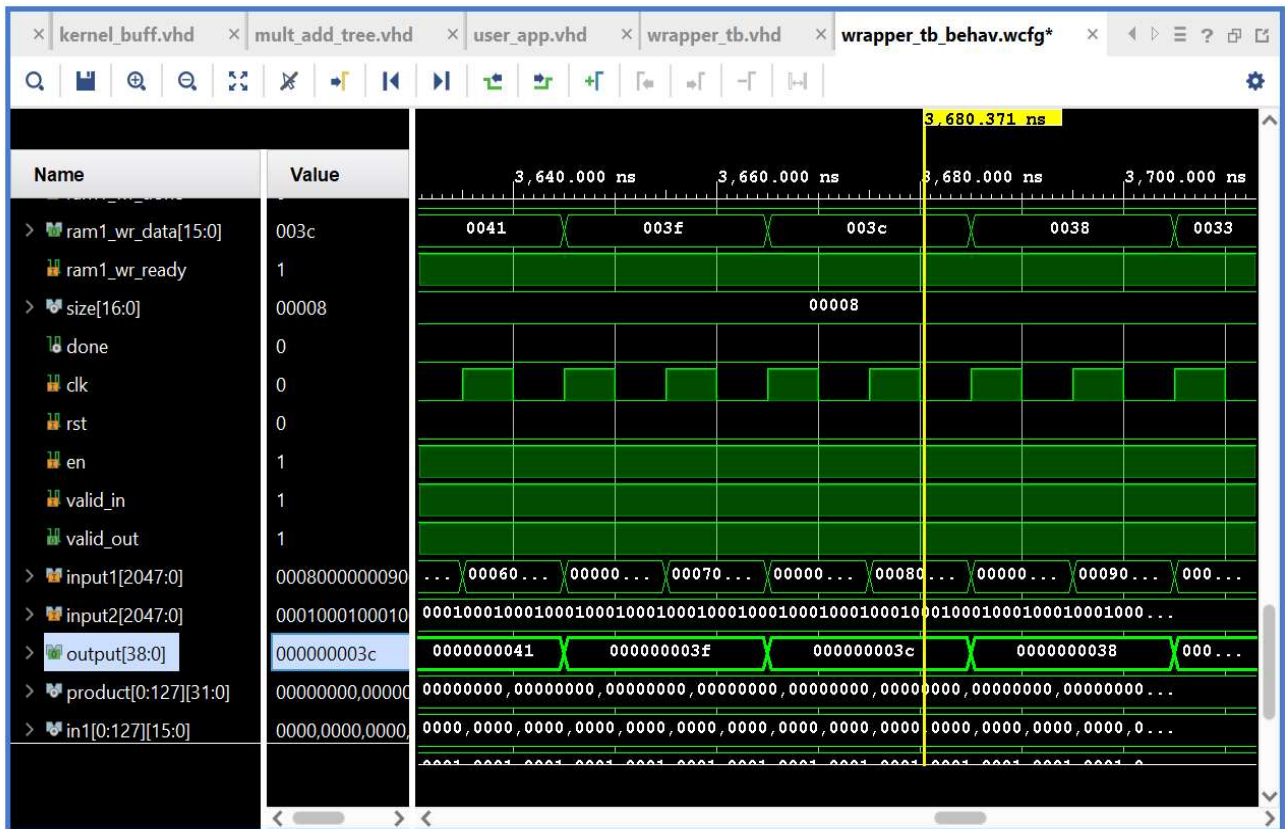


Fig 1: Datapath clipping logic

2. SIGNAL BUFFER:

The input to this entity comes from RAM0. This consists of 128 16-bit registers. In every cycle, the elements will be shifted and one new input arrives and thus has the sliding window ability which is advantageous for convolution as it has huge overlaps. Count signal is added to keep a count of number of elements. We have given Full and Empty flags such that the signal buffer is full whenever there are 128 elements in the registers and it is empty when count < 128.

We made sure to assert the signal buffer wr_en only when RAM1 rd_en is asserted that is when there is valid output from RAM0 and signal buffer is not full. Also, rd_en is asserted only when ram1 is ready, signal buffer is not empty and kernel is loaded.

Signal_rd_en \leftarrow (not Signal_Empty) and ram1_wr_ready and kernel_loaded

ram0_rd_en_signal \leftarrow ram0_rd_valid and (not Signal_Full)

The output of the signal buffer then serves as one of the inputs to convolution pipeline.

For test size : 8 ;

The following screenshots below show the shifting of inputs in the registers :

- (i) Input 0001 is stored in register 0.

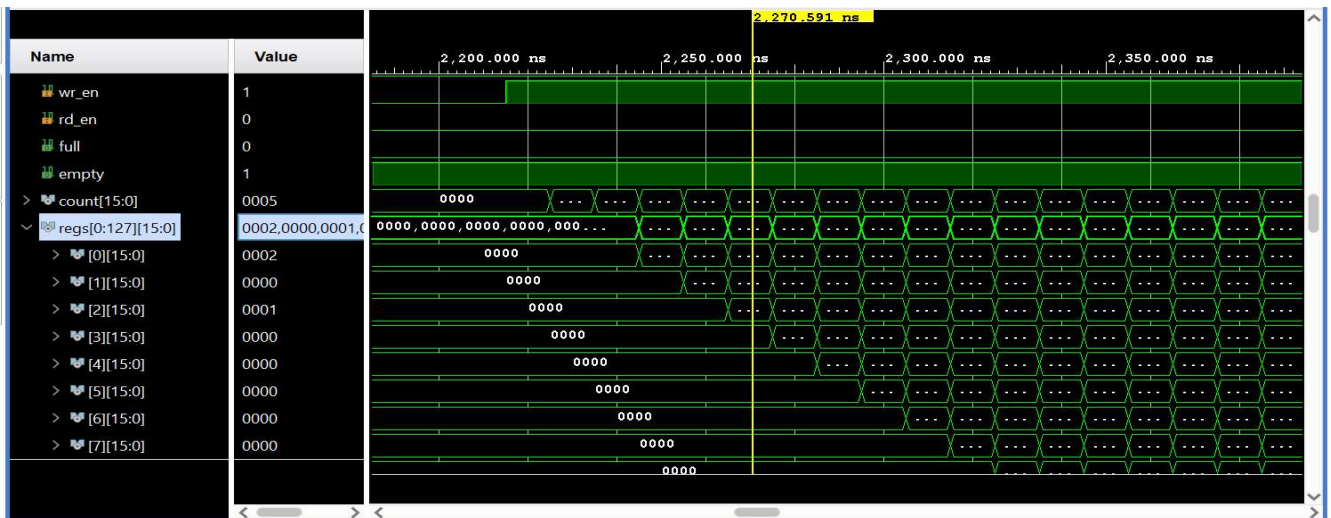


Fig 4: Signal Buffer functioning

- (iv) Previous inputs are shifted and new element 0003 is stored. This goes on till 000b (12) and then padding occurs.

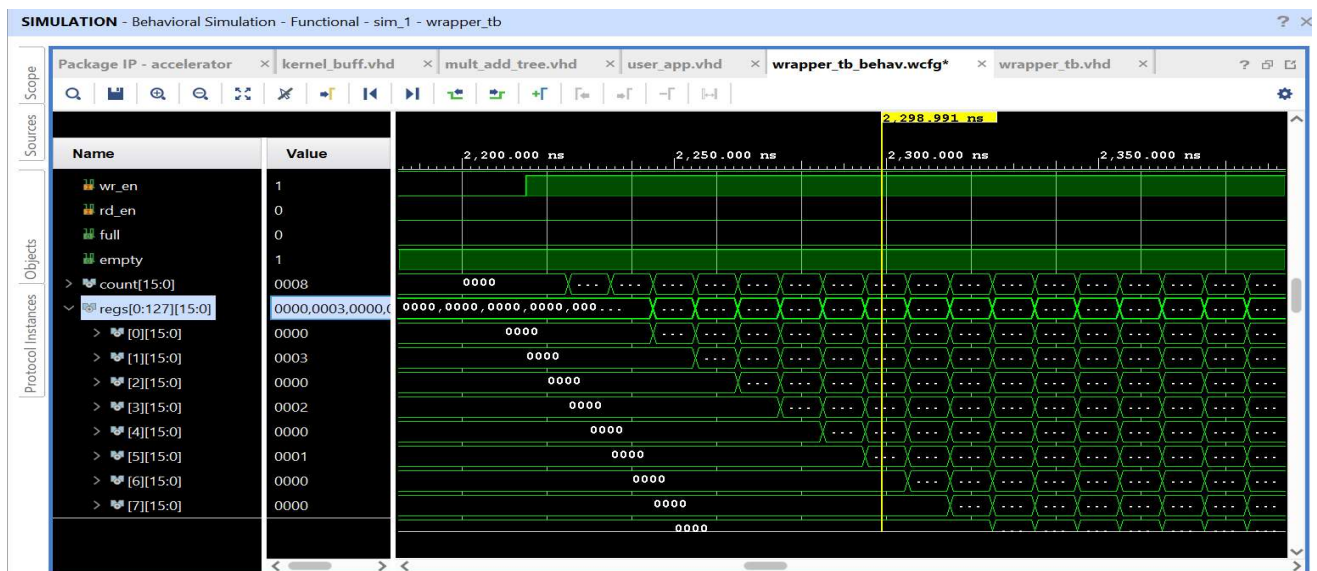


Fig 5: Signal Buffer functioning

The following screenshots show the output of the signal buffer :

- (i) The outputs will be read as shown below from 0001 to 000b along with 0's

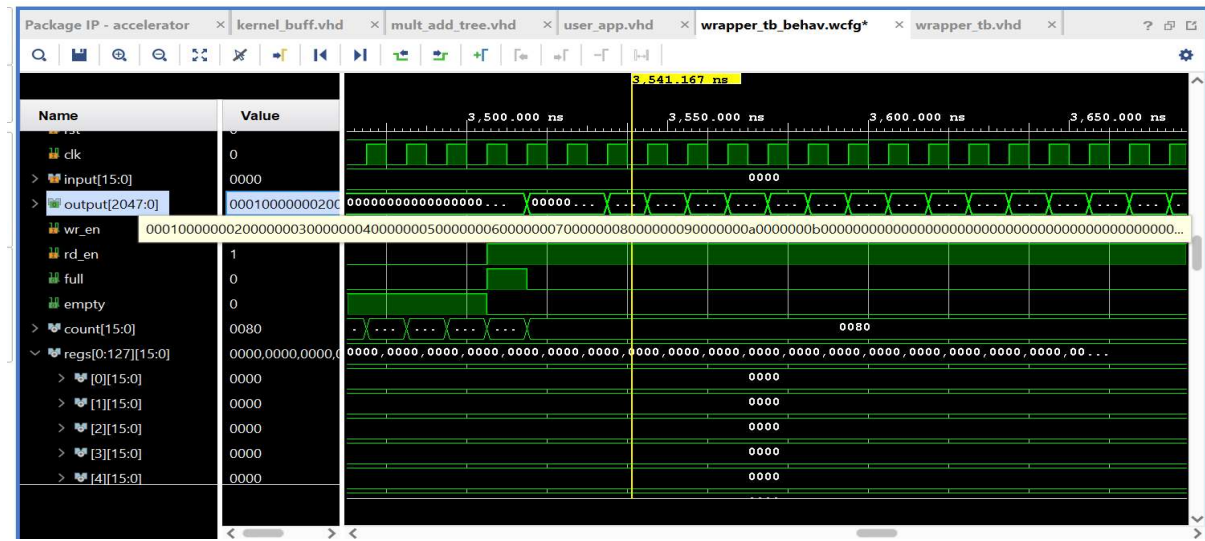


Fig 6: Signal Buffer Output

- (ii) Then in the next step 0001 is discarded. In the later step, 0002 will be discarded and so on.

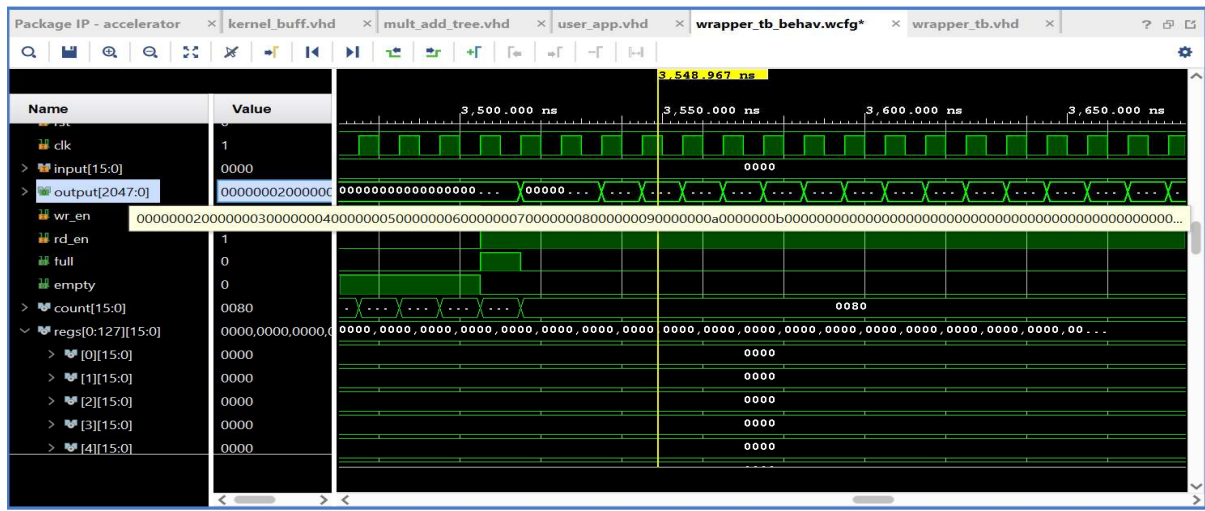


Fig 7: Signal Buffer Output

3. KERNEL BUFFER

The input to this entity comes from memory map. This also consists of 128 16-bit registers. The elements are shifted and new inputs arrive till the count becomes 128. The difference between signal buffer and kernel is that once the inputs are stored in 128 registers, it won't accept or need new elements. In other words, kernel doesn't have sliding window functionality. We designed in such a way that kernel full flag is asserted when count = 128, i.e., there are 128 elements stored in registers otherwise it is not full.

The output of the kernel buffer further serves as another input to convolution pipeline.

The following screenshot shows the input and output of kernel buffer :

Inputs are stored and gets shifted one by one till all the 128 registers are filled.

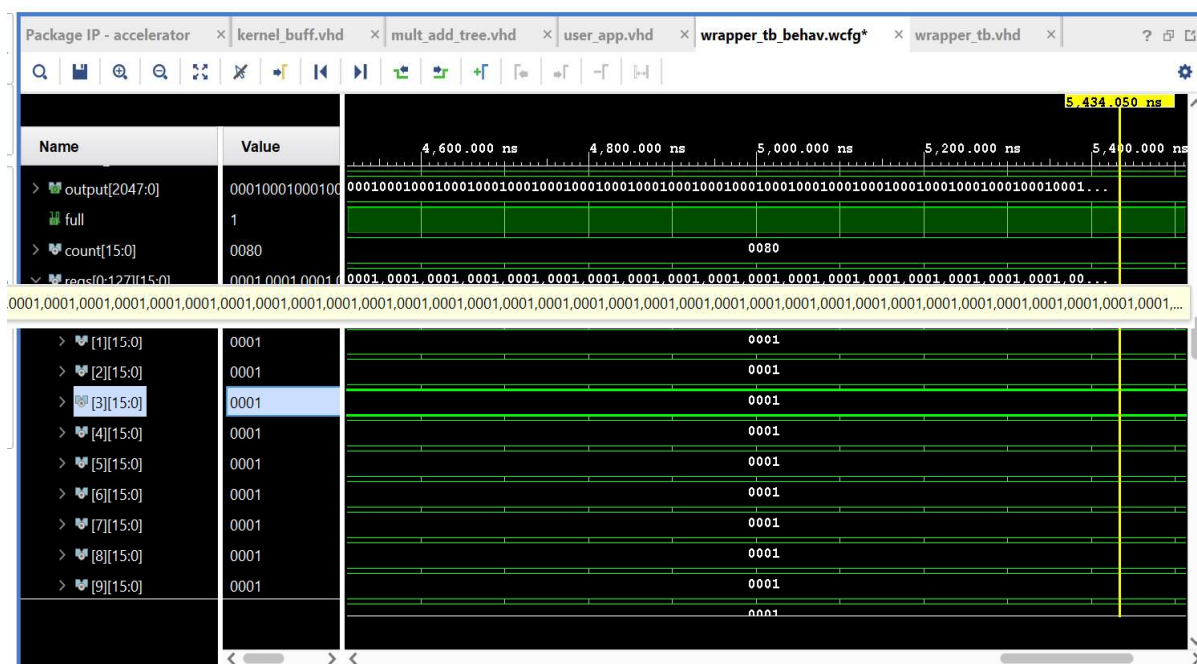


Fig 8: Kernel Buffer

III. RESULTS:

A. DMA INTERFACE

The DMA interface is fully functional on both simulation and FPGA for all test cases.

Hardware output:

```

ptamhankar@ece-b312-recon:~/drum_test
login as: ptamhankar
ptamhankar@reconfig.hcs.ufl.edu's password:
Access denied
ptamhankar@reconfig.hcs.ufl.edu's password:
Last failed login: Wed Dec 15 16:20:39 EST 2021 from 10.228.2.201 on ssh:notty
There was 1 failed login attempt since the last successful login.
Last login: Wed Dec 15 16:16:19 2021 from 10.228.2.201
[ptamhankar@ece-b312-recon ~]$ cd dram_test
[ptamhankar@ece-b312-recon dram_test]$ zed_schedule.py ./zed_app dram_test.bit
Searching for available board....
Starting job "/zed_app dram_test.bit" on board 192.168.1.110:
Programming FPGA....SUCCESS
Testing transfers to/from address 0....SUCCESS
Testing max transfer size....SUCCESS
Testing random sizes and addresses....SUCCESS
[ptamhankar@ece-b312-recon dram_test]$ zed_schedule.py ./zed_app dram_test.bit
Searching for available board....
Starting job "/zed_app dram_test.bit" on board 192.168.1.106:
Programming FPGA....SUCCESS
Testing transfers to/from address 0....SUCCESS
Testing max transfer size....SUCCESS
Testing random sizes and addresses....SUCCESS
[ptamhankar@ece-b312-recon dram_test]$ zed_schedule.py ./zed_app dram_test.bit
Searching for available board....
Starting job "/zed_app dram_test.bit" on board 192.168.1.101:
Programming FPGA....SUCCESS
Testing transfers to/from address 0....SUCCESS
Testing max transfer size....SUCCESS
Testing random sizes and addresses....SUCCESS
[ptamhankar@ece-b312-recon dram_test]$ zed_schedule.py ./zed_app dram_test.bit
Searching for available board....
Starting job "/zed_app dram_test.bit" on board 192.168.1.101:
Programming FPGA....SUCCESS
Testing transfers to/from address 0....SUCCESS
Testing max transfer size....SUCCESS
Testing random sizes and addresses....SUCCESS
[ptamhankar@ece-b312-recon dram_test]$

```

Fig 9: Hardware output for DMA interface

Simulation results for different test sizes and start addresses :

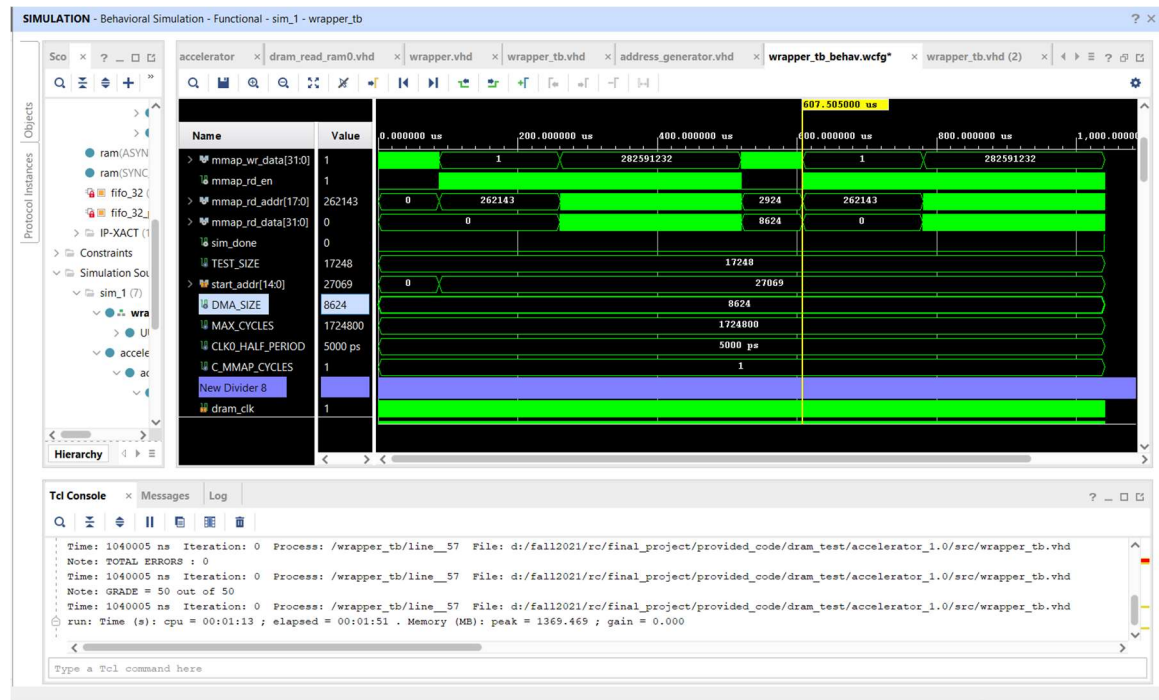


Fig 10: Simulation output for DMA interface with test size 17248 and address 27069

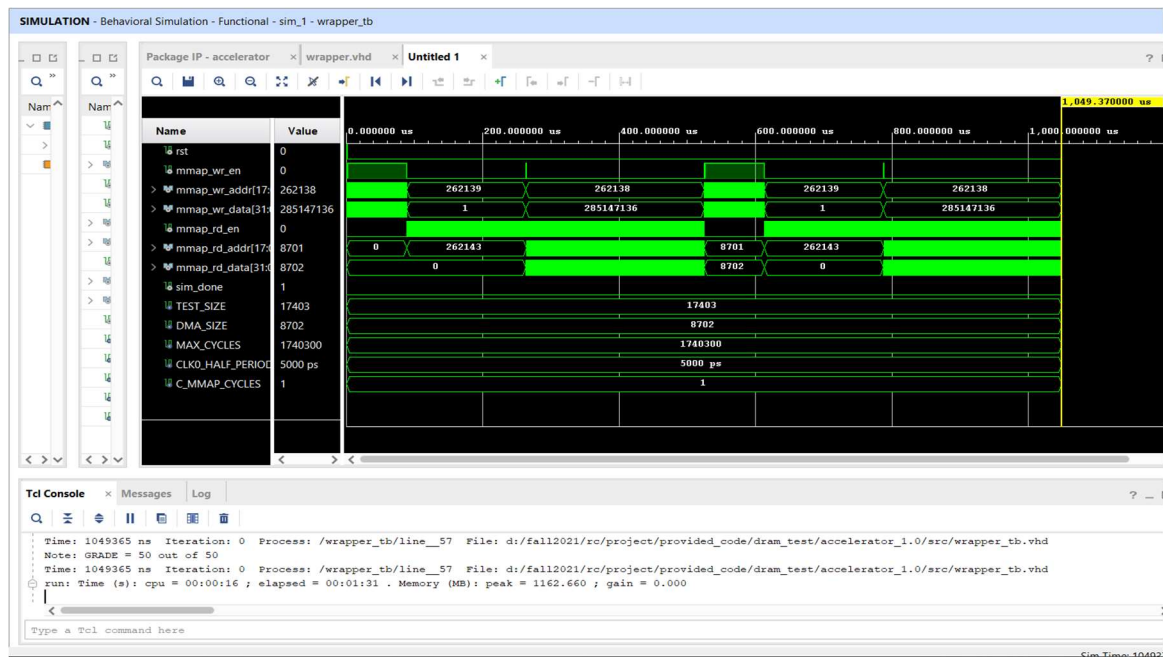


Fig 11: Simulation output for DMA interface with test size 17403 and address 6509

B. USERAPP

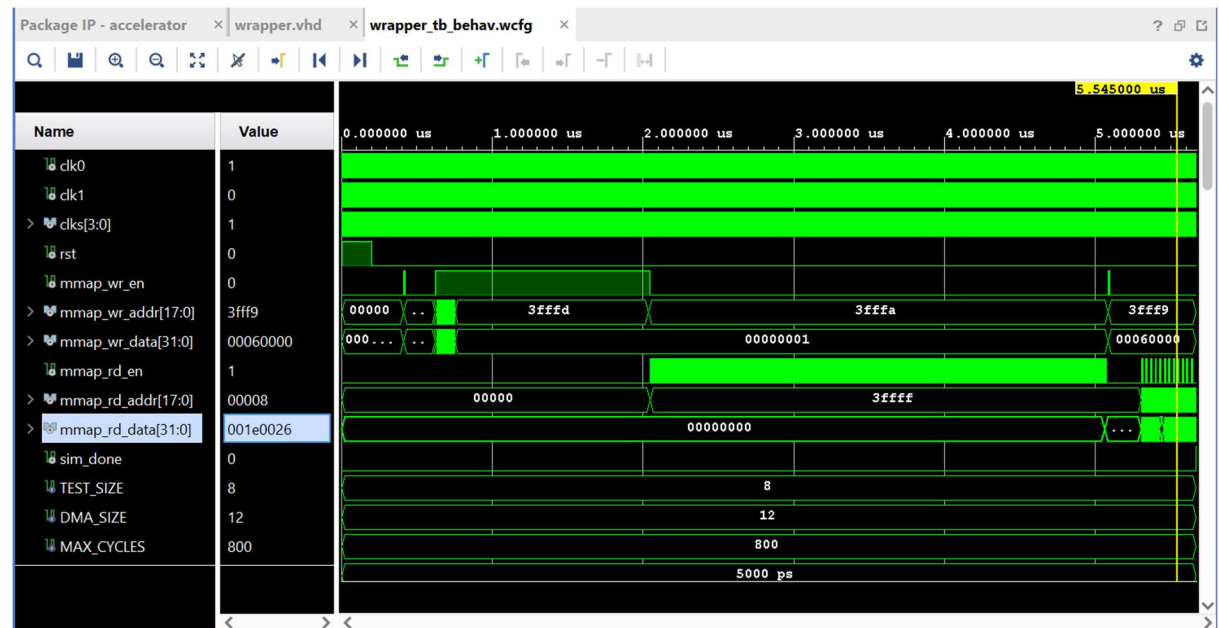


Fig 123: Simulation output for USER_APP for TEST SIZE 8

IV. CHALLENGES:

A. DMA INTERFACE

1. FIFO asserting EMPTY, FULL AND PROGRAMMABLE FULL together
Solution: ANDing the done and go signals solved the reset behaviour of the FIFO IP from the IP catalogue provided by Vivado.
2. FIFO reading inverted bits
Solution: The FIFO was inverting the two halves of the 32 bit word while reading from external memory. By flipping the two halves of the 32 bit word, correct data was obtained in the buffer

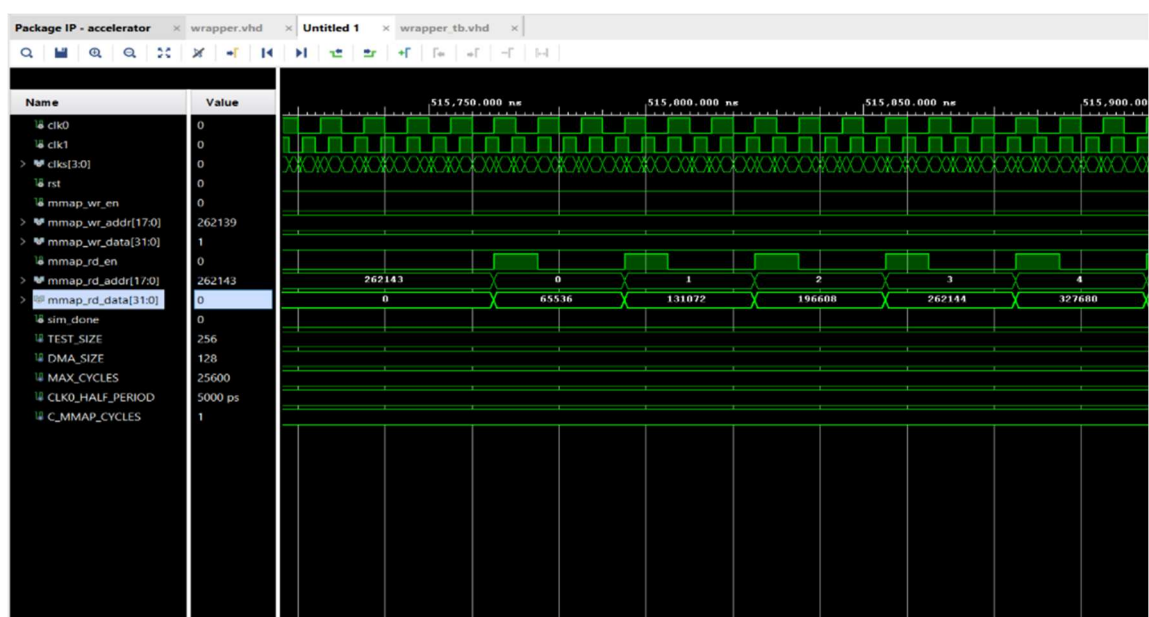


Fig 13: Garbage values in DIN from FIFO due to inverted bits

3. Metastability

Solution: The mistake that was done was adding 3 registers in the source and destination domain which did not solve metastability. After adding one register in source domain and two registers in destination domain for send_s signal and one register in destination domain and two registers in source domain for ack_s signal, metastability was fixed and done was asserted on time

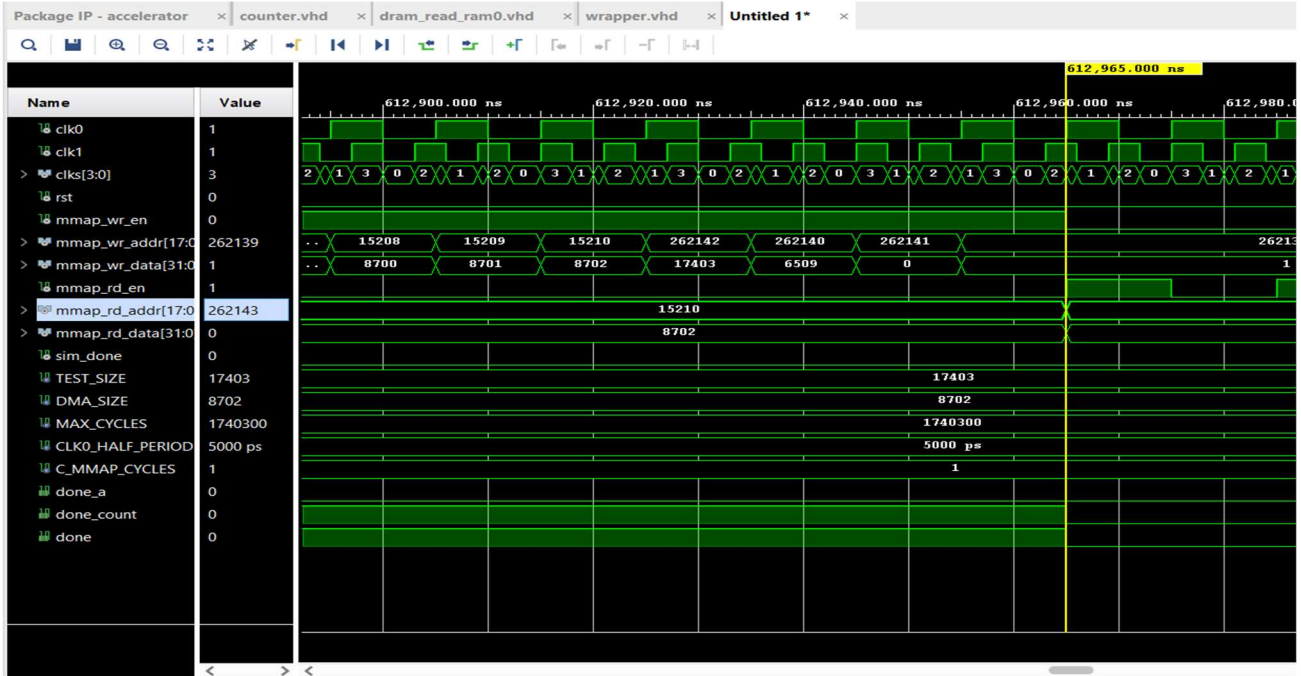


Fig 14: DONE perfectly asserted on simulation for failed test case on board

```
[ptamhankar@ece-b312-recon dma]$ zed_schedule.py ./zed_app dma.bit
Searching for available board....
Starting job "./zed_app dma.bit" on board 192.168.1.108:
Programming FPGA....SUCCESS
Testing transfers to/from address 0....SUCCESS
Testing max transfer size....SUCCESS
Testing random sizes and addresses....(Size = 17403, addr = 6509)Error: Done was
not asserted before timeout.
terminate called after throwing an instance of 'TimeoutException'
what(): Timeout Exception
sh: line 1: 1237 Aborted                  ./zed_app dma.bit
[ptamhankar@ece-b312-recon dma]$ zed_schedule.py ./zed_app dma.bit
Searching for available board....
Starting job "./zed_app dma.bit" on board 192.168.1.108:
Programming FPGA....SUCCESS
Testing transfers to/from address 0....SUCCESS
Testing max transfer size....SUCCESS
Testing random sizes and addresses....(Size = 17403, addr = 6509)Error: Done was
not asserted before timeout.
terminate called after throwing an instance of 'TimeoutException'
what(): Timeout Exception
sh: line 1: 1256 Aborted                  ./zed_app dma.bit
[ptamhankar@ece-b312-recon dma]$ s
```

Fig 15: Done not being asserted on board hinting towards metastability

4. Address Generator giving timing issues

Solution: Converting it into a FSMD solved most of the timing and logical issues that the previous design (behavioural) was causing.

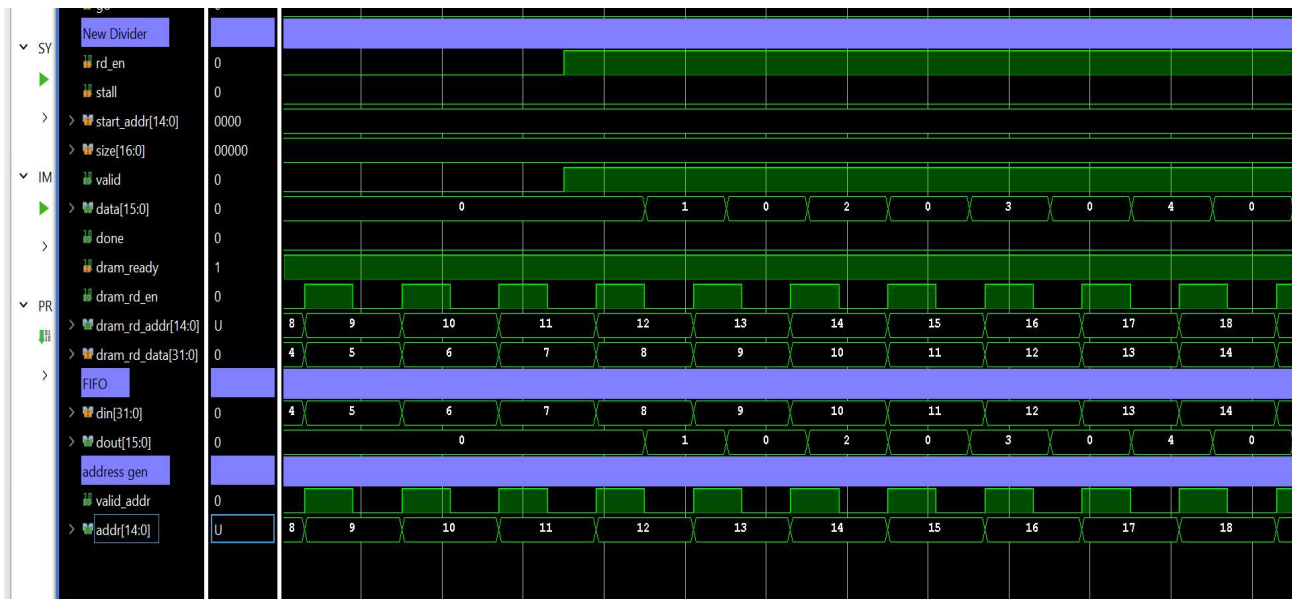


Fig 16: Correct address generator and FIFO functionality

B. USERAPP

- (1) When we designed the signal and kernel buffer, we put them in same directions which caused a problem. We later rectified it by reverting signal buffer to make sure they are in opposite directions while performing convolution.

When signal buffer and kernel are in same directions, we got the performance was very low (39.1066)

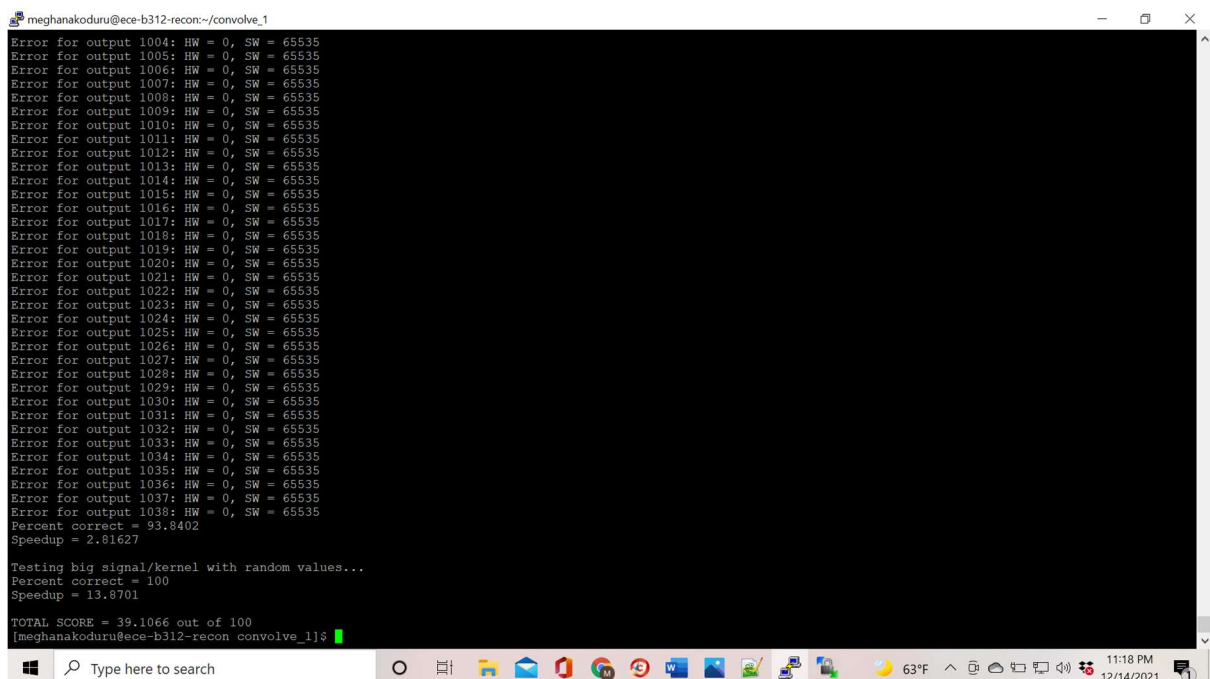


Fig 17: Previous results before invertin

After rectifying the problem by reverting signal buffer, the performance extensively increased to 98.4469

```
meghanakoduru@ece-b312-recon:~/convolve_1$ cd convolve_1
meghanakoduru@ece-b312-recon convolve_1$ make
arm-linux-g++ -o zed_app main.o Board.o Timer.o App.o Convolve.o
meghanakoduru@ece-b312-recon convolve_1$ zed_schedule.py ./zed_app convolve_6.bit
Searching for available board....
Starting job "/zed_app convolve_6.bit" on board 192.168.1.104:
Programming FPGA....Testing small signal/kernel with all 0s...
Percent correct = 100
Speedup = 0.019802

Testing small signal/kernel with all 1s...
Error for output 0: HW = 2, SW = 1
Percent correct = 92.3077
Speedup = 0.0196079

Testing small signal/kernel with random values (no clipping)...
Error for output 0: HW = 93, SW = 26
Percent correct = 92.3077
Speedup = 0.0206186

Testing medium signal/kernel with random values (no clipping)...
Error for output 0: HW = 92, SW = 96
Percent correct = 99.9038
Speedup = 2.69139

Testing big signal/kernel with random values (no clipping)...
Error for output 0: HW = 126, SW = 99
Percent correct = 99.9985
Speedup = 15.9451

Testing small signal/kernel with random values...
Percent correct = 100
Speedup = 0.0181818

Testing medium signal/kernel with random values...
Percent correct = 100
Speedup = 2.33488

Testing big signal/kernel with random values...
Percent correct = 100
Speedup = 14.5546

TOTAL SCORE = 98.4469 out of 100
meghanakoduru@ece-b312-recon convolve_1$
```

Fig 18: Results after inverting

The following screenshots below shows the reversal of signal buffer :

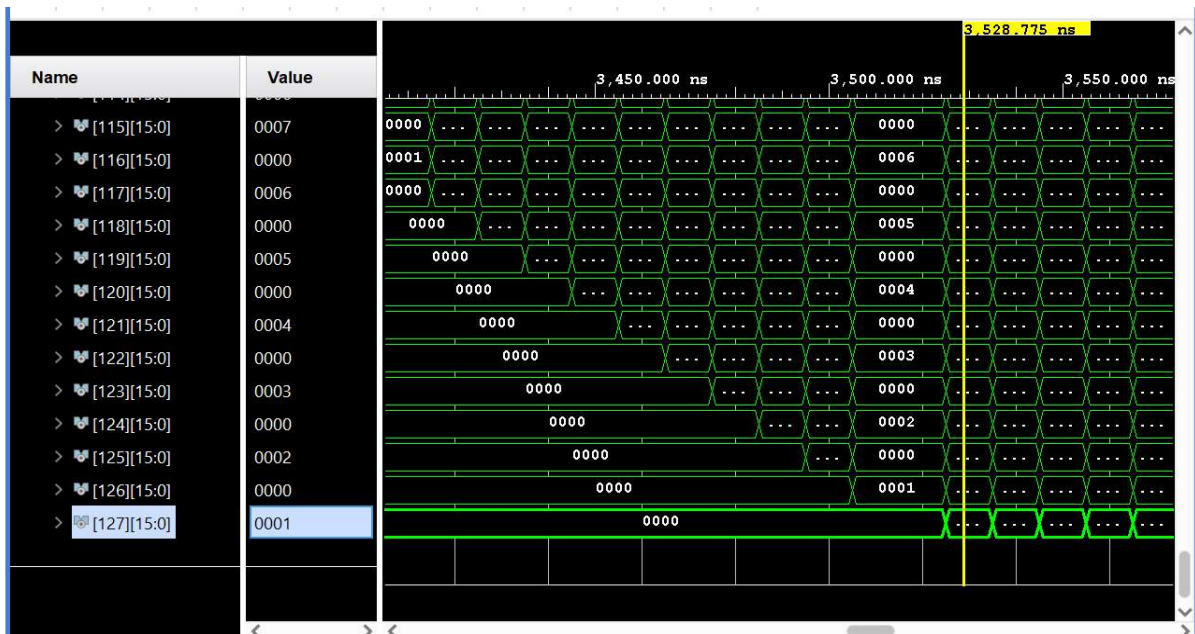


Fig 19: Signal Buffer Inversion

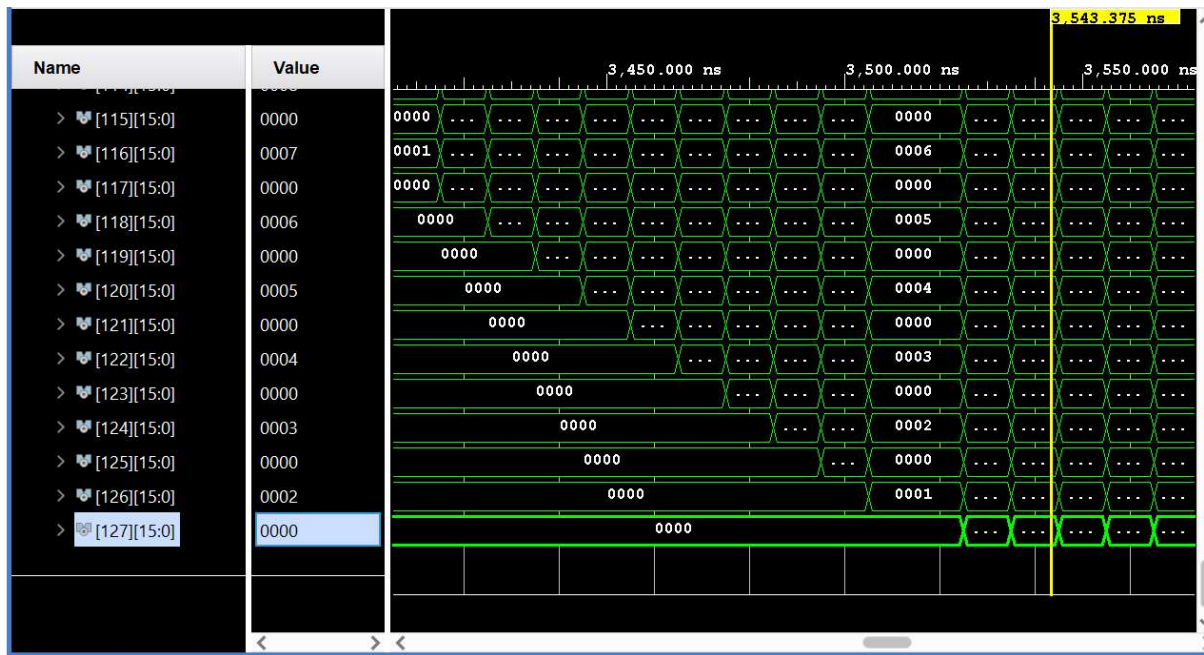


Fig 20: Signal Buffer inversion

- (2) Had an issue while simulating user_app as we forgot to add ‘_0’ to the dram_funcsim files as per instructions given.

V. CONCLUSION:

The project helped gain knowledge about different concepts of digital design and FPGA programming. The concept of design the circuit then write the code definitely helped simplify the design process and writing code to minimize design errors. The DMA interface dealt mainly with handling metastability and address generation. Resolving the timing issues was the main challenge when it came to making the DMA interface work perfectly. For the convolution part, although the score is 98.4469, the signal buffer, kernel buffer and pipeline entities are functioning correctly, when checked individually.