Application Acceleration with High-Level Synthesis

LabC DSP

石思宇 110061422 邱崇喆 108061173

Overview:

The DSP Library implements a discrete Fourier transform using an FFT algorithm for acceleration on AMD Xilinx FPGAs, and it provides a fully synthesizable PL-based SSR FFT as well as a 2-dimensional FFT version.

This library currently supports two data types, fixed point and floating point complex inputs for synthesis.

Define the parameters in FFT structure:

FFT structure is defined as follows:

```
struct ssr_fft_fix_params:ssr_fft_default_params
{
    static const int N = 1024;
    static const int R = 4;
    static const scaling_mode_enum scaling_mode = SSR_FFT_NO_SCALING;
    static const fft_output_order_enum output_data_order = SSR_FFT_NATURAL;
    static const int twiddle_table_word_length = 18;
    static const int twiddle_table_integer_part_length = 2;
    static const transform_direction_enum transform_direction = FORWARD_TRANSFORM;
    static const butterfly_rnd_mode_enum butterfly_rnd_mode = TRN;
};
```

fig.1 FFT structure

In reference to the information in

https://xilinx.github.io/Vitis Libraries/dsp/2019.2/user guide/L1.html:

- N is the size of transform
- R is the number of samples to be processed in parallel SSR Factor and radix of FFT algorithm used
- scaling_mode: The scaling mode as enumeration type (SSR FFT has three different scaling modes)
- output_data_order: Which will decide if data will be in natural order or digit reversed transposed order

- twiddle_table_word_length: Defines total number of bits to be used for storing twiddle table factors
- twiddle_table_integer_part_length: The number of integer bits used for storing integer part of twiddles
- transform_direction : Defines the direction of transform, inverse transform
 (SSR IFFT) or forward transform (SSR FFT)
- butterfly_rnd_mode : Defines the rounding mode used by butterflies in SSR
 FFT stages

Data types:

As mentioned earlier, currently the FFT supports fixed point and floating point complex input for synthesis, where fixed point data type should be declared as std::complex<ap_fixed<>> for both synthesis and simulation. While one can use floating point types std::complex<float> and std::complex<double> for simulation, they are not synthesizable because they might lead to additional resources — more DSP blocks. Instead, we can use complex_wrapper<float> for wrapping complex float data and synthesis.

Туре	Synthesis	Simulation	
std::complex <ap_fixed <="">></ap_fixed>	YES	YES	
std::complex <float></float>	NO	YES	
std::complex <double></double>	NO	YES	
complex_wrapper <double></double>	NO	YES	
complex_wrapper <float></float>	YES	YES	

fig.2 table indicating which data types are synthesizable, extracted from https://xilinx.github.io/Vitis_Libraries/dsp/2019.2/user_guide/L1.html:

1D fixed point FFT with impulse test data:

Test program & test data (The red segment creates impulse data):

How to run?

- 1. source /opt/Xilinx/Vitis/2022.1/settings64.sh
- 2. export

DEVICE=/opt/xilinx/platforms/xilinx_u50_gen3x16_xdma_5_202210_1/xilinx_u50_gen3x16_xdma_5_202210_1.xpfm

3. modify the Makefile as the following:

```
185 # Alias to run, for legacy test script
186 check: run
187
188 CSIM ?= 1
189 CSYNTH ?= 1
190 COSIM ?= 1
191 VIVADO_SYN ?= 0
192 VIVADO_IMPL ?= 0
193 QOR CHECK ?= 0
194
195 # at least RTL synthesis before check QoR
196 ifeq (1,$(QOR_CHECK))
197 ifeq (0,$(VIVADO_IMPL))
198 override VIVADO_SYN := 1
199 endif
200 endif
201
202 # need synthesis before cosim or vivado
203 ifeq (1,$(VIVADO_IMPL))
204 override CSYNTH := 1
205 endif
206
207 ifeq (1,$(VIVADO_SYN))
208 override CSYNTH := 1
209 endif
210
211 ifeq (1,$(COSIM))
212 override CSYNTH := 1
213 endif
```

4. make run

We'll can see the project have been created named proj_impulse_test.prj, inside which we can find the the reports generated by vitis hls:

synthesis:

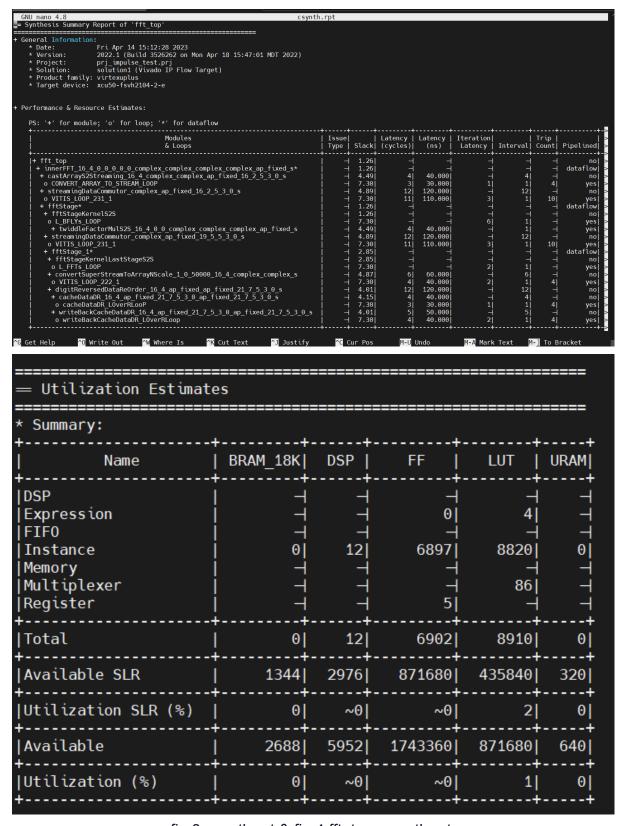


fig.3 csynth.rpt & fig.4 fft_top_csynth.rpt

c-sim:

fig.5 c-simulation result

Note: There was nothing printed on the log file, which means the result was correct, because in main.cpp we added a "cout" to ensure the output of an impulse input is one:

fig.6 testbench (main.cpp) modification

co-sim:

fig.7 co-sim result

Multiple instances:

We can create multiple instance of SSR FFT by changing the second template integer parameter:

```
xf::dsp::fft::fft<fftParams,1><ssr_fft_fix_params>(...);
xf::dsp::fft::fft<fftParams,2><ssr_fft_fix_params>(...);
```

The following showcase a simple code for creating one more instance to run FFT of constant one (or the inverse FFT of the previous result), by modifying the main.cpp and top_module :

fig.8 modified version of tb for testing multiple instance

```
27 // Define FFT Size and Super Sample Rate
28 #define FFT_LEN 16
29 #define SSR 4
30
31 #define IN_WL 16
32 #define IN_IL 2
33 #define TW_WL 16
34 #define TW_IL 2
35 #define IID 0
36 #define IID_NEW 1
```

fig.9 modification of top function — add an instance

we can see every resource in utilization estimates of the top function were nearly doubled, because we instantiated another one.

* Summary:										
+ Name	++ BRAM_18K	DSP	FF	LUT	URAM					
DSP Expression FIFO Instance Memory Multiplexer Register		- - - 28 - -	⊣ 0 ⊣ 13806 ⊣ ⊣ 8	- 10 - 18224 - 158	⊣ 0 ⊣					
Total	0	28	13814	18392	0					
Available SLR	1344	2976	871680	435840	320					
Utilization SLR (%)	0	~0	1	4	0					
Available	2688	5952	1743360	871680	640					
Utilization (%)	++ 0	 ~0	 0~	2	+ 0					

fig.9 utilization estimates of creating two instances

and the c-sim and co-sim were also successfully run:

fig.10 successful c-sim & co-sim of creating two instances

1D float point FFT with test data in /L1/tests/common_float/verif/:

The test data input file used in this experiment is fftStimulusIn_L4096.verif, and the golden output file used for verification is fftGoldenOut L4096.verif:

```
2 9.20659596040256644756
 3 12.92746544283520648833
 4 5.93800120776925144384
 5 -2.86931512498761431829
 6 9.91246943611761288651
    -2.66690187629595776286
8 -17.63462151916279907482
9 -19.95775095737172577515
10 -31.32222132136553938153
11 -3.08521101278948517432
12 -7.79760667271630225628
13 -5.79936656748928491822
14 7.56072172358180694118
    -7.09100249994199849368
    -24.50207072565649468743
17
    -1.76773374266216953821
18 2.11522636081379555861
19 -3.00672079096654876196
20 26.42191696659537214487
21 -8.13999728384190035513
22 -0.81171396629295222702
23 -40.51150184918373753362
24 0.33088126730945377485
25 16.97813990267743378126
    17.96349704351164788818
27 -9.87015850993069676633
28 -12.27577156099332356121
29 -14.76402930727036988401
30 1.63618061082191768030
31 -39.62788907700327456496
```

fig.11 a snippet of test data in fftStimulusIn L4096.verif

synthesis summary:

c-sim:

co-sim:

	Report time : Sat 15 Apr 2023 02:37:14 PM CST. Solution : solution1. Simulation tool : xsim.									
<u> </u>	 RTL +	Status +	Laten min	avg	max	In min	terval(Clock Cycl avg	.es) max	Total Execution Time - (Clock Cycles) + 	
-	VHDL Verilog	NA Pass	NA 2497	NA 2497	NA 2497	NA NA	NA NA	NA NA	NA 2497 	

2D fixed point FFT with impulse test data:

test program and test data:

```
printMatStream<k_fftKernelSize, k_fftKernelSize, k_memWidth, MemWideIFTypeOut>(
fftOutputStream, "2D FFT Output Natural Order...");
streamTOMatrix<k_fftKernelSize, k_fftKernelSize, k_memWidth, T_outType>(fftOutputStream, l_outMat);
/ runs loop
                                 std::cout « "-----" « std::endl; std::cout « "------" ( std::endl; std::cout « "--------" ( std::endl; std::cout « "--------" ( std::endl; std::cout « "-------" ( std::endl; st
    72 #endif
```

Similar FFT structure definition for 2D FFT can be found in top 2d fft test.hpp:

```
46 struct FFTParams : ssr_fft_default_params {
47     static const int N = k_fftKernelSize;
48     static const int R = k_fftKernelRadix;
49
50
         static const scaling_mode_enum scaling_mode = SSR_FFT_NO_SCALING;
51
52 }
         static const transform_direction_enum transform_direction = FORWARD_TRANSFORM;
```

synthesis:

c-sim:

```
242
243
244
245
246 2D FFT Output Natural Order...
247 Stream Size :32
248 (1,1) (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1), (1,1)
```

co-sim:

Report time Solution Simulation to										
 + RTL + 	 Status 	Lat min	ency(Clock Cycles avg	max	In min	terval(Clock Cyc avg	les) max	Total Execution Time + (Clock Cycles) + 		
VHDL Verilog ++	NA Pass	NA 1470	NA 1470	NA 1470	NA 1471	NA 1471	NA 1471	NA 7354		

2D float point FFT with test data in

~//LAB_C/Vitis_Libraries/dsp/L1/tests/common_2dfft/2dFFTVerificationData/."

The test data input file used in this experiment is fft2DStimulusIn_L256.verif, and the golden output file used for verification is fft2DGoldenOut L256.verif:

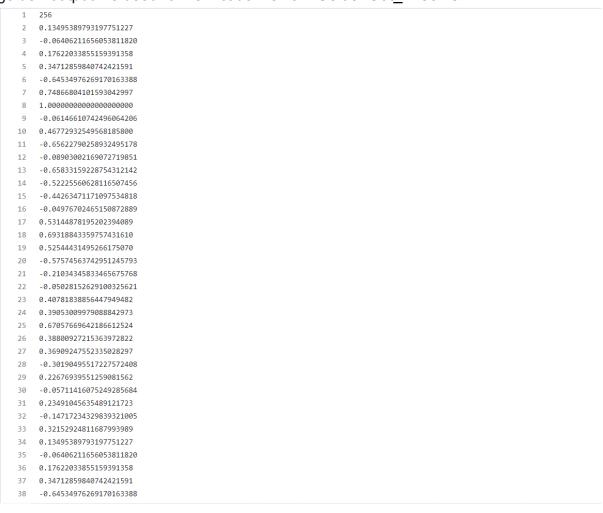


fig.12 a snippet of test data in fft2DStimulusIn_L256.verif

test program (red segment part loads the input file):

synthesis summary:

c-sim:

```
The Main Memory witch (me. complex/floats) : 6

The Size of ID Row Kernel : 4

The Size of ID Row Kernel : 4

The Transform Direction for Row Kernel : 16

The Size of ID Column Kernel : 16

T
```

co-sim:

Report time : Mon 17 Apr 2023 03:42:45 AM CST. Solution : solution1. Simulation tool : xsim.										
- RTL	 + Status 	 min	Lateno 	y(Clock Cycle avg	s) max	In min	terval(Clock Cyc avg	les) max	Total Execution Time - (Clock Cycles) + 	
VHDL Verilog			NA 2342	NA 3743	NA 4094	NA 591	NA 1905	NA 2344	NA 11714 	

L2 of SSR FFT

As the L3 function for DSP library hasn't been included in the project, our principal goal in the second part is to identify L2 modules, use the kernel's appropriate sources to build XCLBIN file to run 1) software emulation, 2) hardware emulation, and 3) build and test on hardware. The kernel for this experiment will be SSR FFT. For this part, we will only cover the openCL kernel as instructed. Following the instruction instruction from the readme file in the library, the files which will be needed to run the OpenCL kernel application project can be found inside the directory "test" under L2.

Just like L1, SSR FFT's L2 consists of 1D and 2D FFT, each in their turn consists of fixed point and floating point data type. Due to their similarity when implementing and constant error from the platform for floating point, we've decided to work only on 1D's fixed point.

Create and Run the project

- 1. Open Vitis IDE
- 2. Select create application project, in Platform windows, choose u50 acceleration card.
- Import the sources needed for both the kernel and the host, these files can be found in 1) L2/tests/hw/1dfft/fixed/host and kernel, 2)
 L2/include/hw/vitis fft/fixed and 3) L1/include/hw/vitis fft/fixed.
- 4. To run the emulations and Hardware test, do as follow for each build configuration:
 - a. build: 1- kernel, 2- hw link, 3-host, 4-system.
 - run configuration with setting: Xilinx xilinx_u50_gen3x16_xdma_5_202210_1 -xclbin
 ./binary_container_1.xclbin, with host trace and openCL trace checked.
 - c. Check run_summary.xrt to visualize and analyze the timeline-trace with Vitis Analyzer.

Host's main code snippet

The following figures show some important part of the host, in Fig 2.1, it shows the declaration and initialization of memory locations that will hold input and output data. Fig 2.2 shows the declaration and how the memory mapping between DDR and the

host's virtual memory is done, using OpenCL's buffer(type of memory object used to store data in the global memory space). Fig 2.3 shows the code line for data transfer between the kernel and global memory.

```
ap_uint<512>* inData = aligned_alloc<ap_uint<512> >(FFT_LEN * nffts / SSR);
ap_uint<512>* outData = aligned_alloc<ap_uint<512> >(FFT_LEN * nffts / SSR);
// impulse as input
for (int n = 0; n < nffts; ++n) {
    for (int t = 0; t < FFT_LEN / SSR; ++t) {
        if (t = 0)
            inData[n * FFT_LEN / SSR + t] = 1;
        else
            inData[n * FFT_LEN / SSR + t] = 0;
}
std::cout << "Host buffer has been allocated and set.\n";</pre>
```

Fig 2.1.

fig 2.2.

```
// write data to DDR
std::vector<cl::Memory> ib;
ib.push_back(in_buff);
q.enqueueMigrateMemObjects(ib, 0, nullptr, &write_events[0][0]);

// set args and enqueue kernel
int j = 0;
kernel.setArg(j++, in_buff);
kernel.setArg(j++, out_buff);
kernel.setArg(j++, nffts);
q.enqueueTask(kernel, &write_events[0], &kernel_events[0][0]);

// read data from DDR
std::vector<cl::Memory> ob;
ob.push_back(out_buff);
q.enqueueMigrateMemObjects(ob, CL_MIGRATE_MEM_OBJECT_HOST, &kernel_events[0], &read_events[0][0]);

// wait all to finish
q.flush();
q.finish();
gettimeofday(&end_time, 0);
```

Fig 2.3.

Software emulation

Allows us to test and debug on a host before developing on FPGA. Below, Fig 2.4 shows software emulation's profile summary while 2.5 and 2.6 show the timeline trace(write, kernel execution, read). The kernel takes 59807.200 ms to execute.

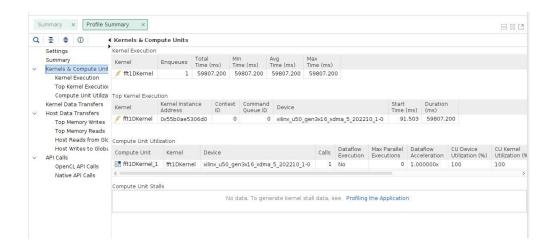


Fig 2.4

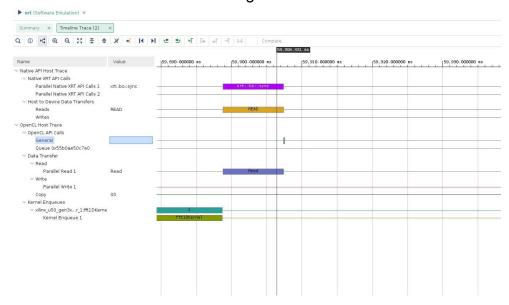


Fig 2.5.

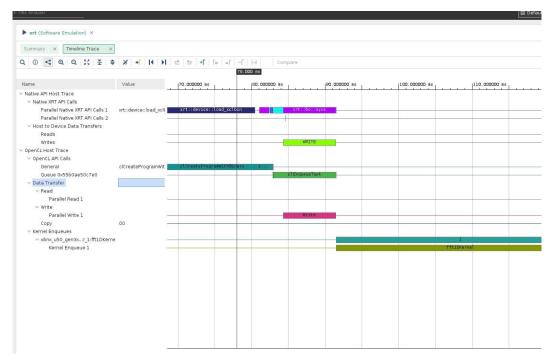


Fig 2.6

Hardware Emulation

Because we were dealing with a large set of data 1024 points FFT, the hardware emulation took longer than the time we had in our hands. Fig 2.7shows the console when executing hardware emulation.

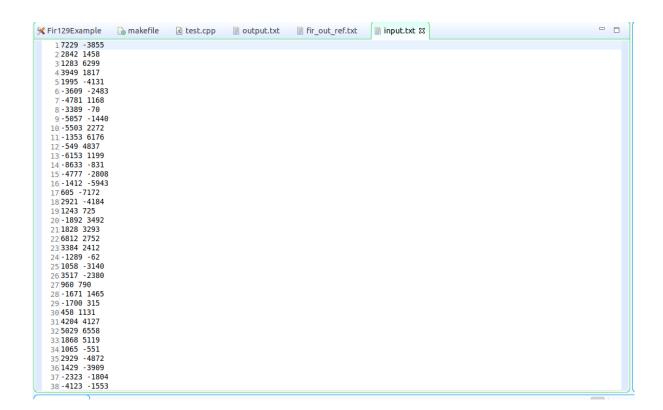
```
Host buffer has been allocated and set.
Found Platform
Platform Mane: Xilinx
Info: Context created
Info: Command queue created
Selected Device xilinx_USO_gen3x16_xdma_5_202210_1
Selected Device xilinx_USO_gen3x16_xdma_5_202210_1
INFO: Importing /mmt/HLSNAS/02.tobflU/Labc/DSP_ldfft_system/Emulation-Hw/binary_container_1.xclbin
Loading: /mmt/HLSNAS/02.tobflU/Labc/DSP_ldfft_system/Emulation-Hw/binary_container_1.xclbin'
INFO: [HW-EMU 01] Hardware emulation runs simulation underneath. Using a large data set will result in long simulation times. It is reco configuring dataflow mode with ert polling
scheduler config ert(1), dataflow(1), slots(16), cudma(0), cuisr(0), cdma(0), cus(1)
INFO: [HW-EMU 07-0] Please refer the path "/mnt/HLSNAS/02.tobflU/Labc/DSP_ldfft/Emulation-Hw/.run/29829/hw_em/deviceO/binary_O/behav_wav
Info: Program created
Info: Kernel created
Kernel created
Kernel created
EMI 07-0] Please refer the path "/mnt/HLSNAS/02.tobflU/Labc/DSP_ldfft/Emulation-Hw/.run/29829/hw_em/deviceO/binary_O/behav_wav
Info: Program created
INFO:: [Vitis-EM 22] [Time elapsed: 4 minute(s) 52 seconds, Emulation time: 0.889473 ms]
Data transfer between kernel(s) and global memory(s)
INFO:: [Vitis-EM 22] [Time elapsed: 9 minute(s) 52 seconds, Emulation time: 1.64863 ms]
Data transfer between kernel(s) and global memory(s)
INFO:: [Vitis-EM 22] [Time elapsed: 14 minute(s) 52 seconds, Emulation time: 2.46379 ms]
Data transfer between kernel(s) and global memory(s)
INFO:: [Vitis-EM 22] [Time elapsed: 19 minute(s) 53 seconds, Emulation time: 2.46379 ms]
Data transfer between kernel(s) and global memory(s)
INFO:: [Vitis-EM 22] [Time elapsed: 24 minute(s) 53 seconds, Emulation time: 3.75205 ms]
Data transfer between kernel(s) and global memory(s)
INFO:: [Vitis-EM 22] [Time elapsed: 24 minute(s) 53 seconds, Emulation time: 5.09732 ms]
Data transfer between kernel(s) and global memory(s)
INFO:: [Vitis-EM 22] [Time elapsed: 24 minute(s) 53 seconds, Emulation time: 5.09732 ms]
Data transfer between kernel(s) and global memory(s)
INFO:: [
```

Fig 2.7

Running fir_129t_sym, an L2 example:

test program & test data:

```
output.txt ☐ fir_out_ref.txt ☐ fir_out_ref.txt ☐ fir_out_ref.txt
                                                                                                                                              input.txt
 13 * limitations under the License.
 15 #pragma once
 16
17 #include "system settings.h"
 19 using namespace adf;
 21⊖ class FirKernel: public graph
 23 private:
24
25
             // FTR coefficients
            std::vector<int16> m taps = std::vector<int16>{-1, -3, 3, -1, -3,
                  ::Vector<intib> m taps = std::Vector<intib</ri>
6, -1, -7, 9, -1, -12, 14, 1, -20, 19, 5, -31, 26, 12, -45, 32, 23, -63, 37, 40, -86, 40, 64, -113, 39, 96, -145, 33, 139, -180, 17, 195, -218, -9, 266, -258, -53, 357, -299, -118, 472, -339, -215, 620, -376, -360, 822, -409, -585, 1118, -437, -973, 1625, -458, -1801, 2810, -470, -5012, 10783, 750273
 26
27
 28
29
 30
31
                  25067};
 32
33
            //FIR Graph class
            //rin Graph class
xf::dsp::dir::sr_sym::fir_sr_sym_graph<DATA_TYPE, COEFF_TYPE, FIR_LEN, SHIFT,
ROUND_MODE, WINDOW_SIZE> firGraph;
 34
35
 36 public:
 38
39
            port<output> out;
             // Constructor - with FIR graph class initialization
 40
            FirKernel():firGraph(m_taps) {
 419
                  // Make connections
// Size of window in Bytes
 42
 43
                   // Jaze or window in bytes.
// Margin gets automatically added within the FIR graph class.
// Margin equals to FIR length rounded up to nearest multiple of 32 Bytes.
 44
                   connect<>(in, firGraph.in[0]);
connect<>(firGraph.out[0], out);
 48
            };
 49 };
```



include file:

fir_sr_sym.cpp fir_sr_asym.cpp widget_api_cast.cpp

L2:

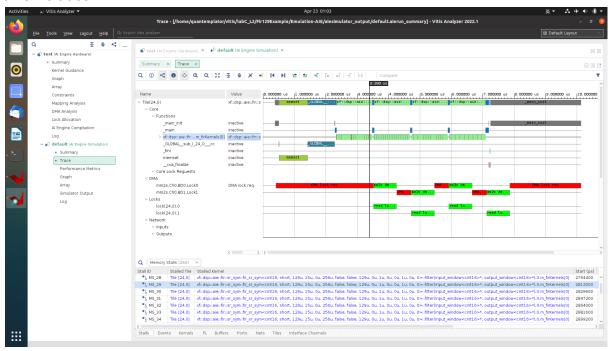
fir_graph_utils.hpp fir_sr_sym_graph.hpp fir_sr_asym_graph.hpp graph_utils.hpp

L1:

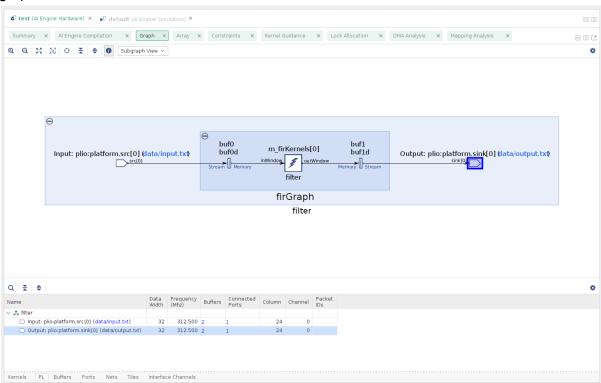
fir_common_traits.hpp
fir_params_default.hpp
fir_sr_asym_traits.hpp
fir_sr_asym_utils.hpp
fir_sr_asym_traits.hpp
fir_sr_sym_traits.hpp
fir_sr_sym_utils.hpp
fir_sr_sym_utils.hpp
fir_sr_sym.hpp
fir_utils.hpp
widget_api_cast_traits.hpp
widget_api_cast_utils.hpp
widget_api_cast.hpp
kernel_api_utils.hpp
kernel_broadcast.hpp
kernel_coeff_reload.hpp

RUN AIE EMULATOR and compare the generated output file with reference output:

timeline trace:



graph:



aie array:



AIE reference documents:

- 1. https://www.xilinx.com/content/dam/xilinx/support/documents/sw_manuals/xilinx2022_2/ug1079-ai-engine-kernel-coding.pdf
- 2. https://xilinx.github.io/Vitis_Libraries/dsp/2021.1/user_guide/L2/3-using-examples.html

Lesson and observation:

During the lab experiment, we were facing lots of problems, such as understanding the file dependencies, and the functions provided in the program files. To find file dependency we had to investigate the run_hls.tcl and the makefile, and there are some needed modifications in the makefile in order to run vitis hls by "make run" command. These things took us plenty of time, but we definitely ended up learning a lot as expected from the professor.

The other part we took some time was the AIE example, since we had to figure out what is AIE and understand the source codes, also the required header files and their dependencies.

github link:

邱崇喆: https://github.com/ccontemplator/AAHLS_LAB_C
石思宇: https://github.com/freud96/AAHL LABC DSP