

TCP Offload Engine PoC Sharing

xu, jia

Sharing

TCP Offload Engine PoC

Xu,Jia



Agenda

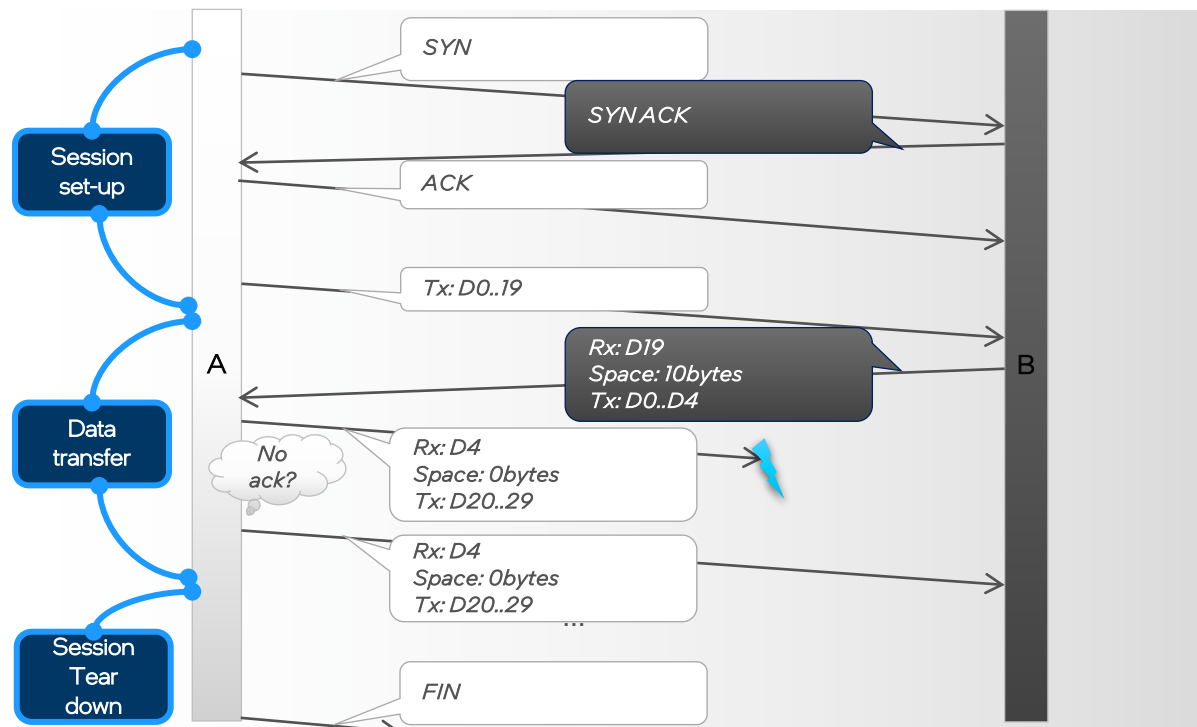
- Overall Architecture
- Conversion Flow
- HLS tuning

Introduction

Subtitle Text Goes Here

Overall Architecture

Introduction



How TCP/IP work

1.Session Setup:

1. Client sends SYN to server.
2. Server responds with SYN-ACK.
3. Client sends ACK, establishing the connection.

2.Data Transfer:

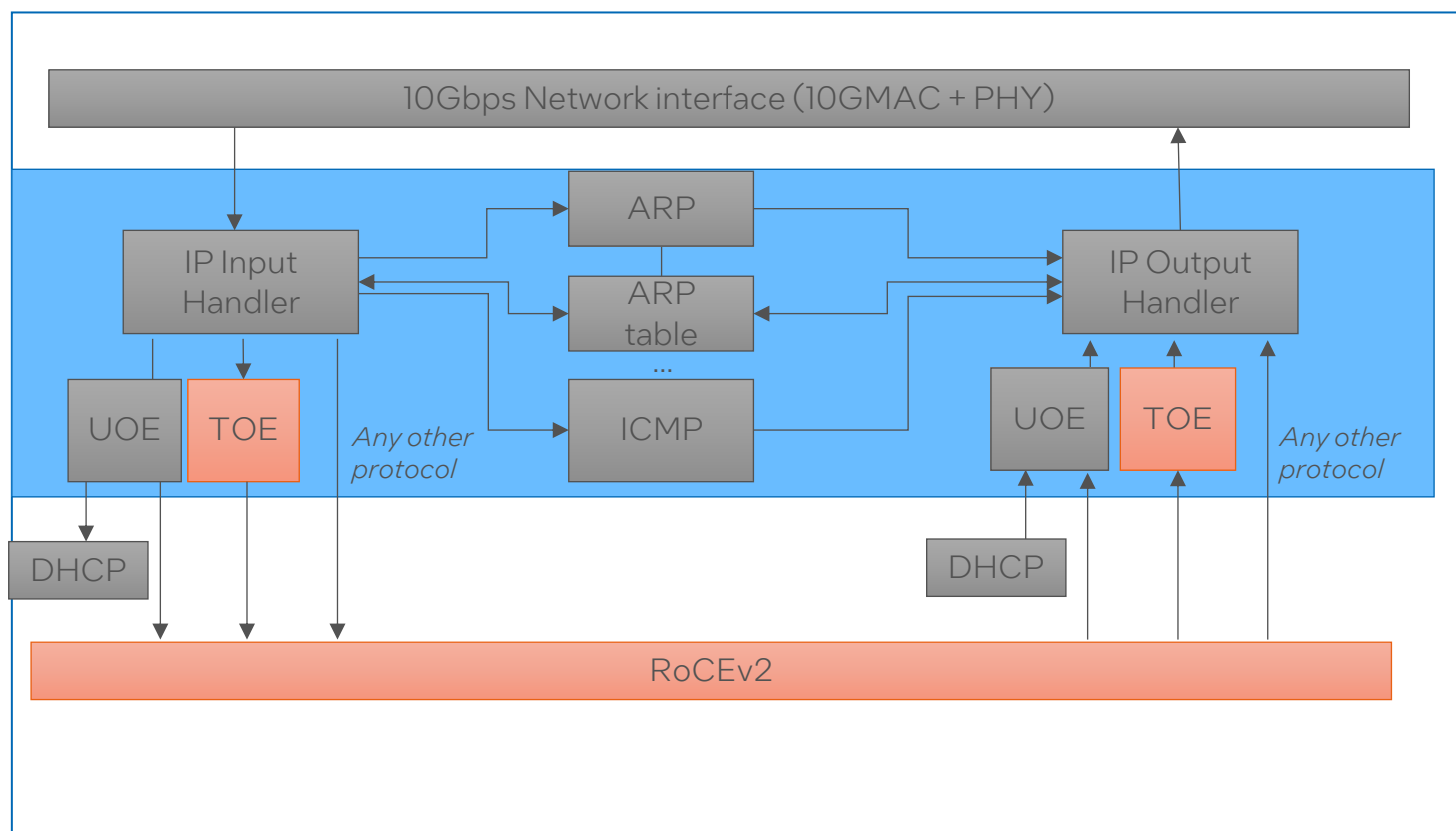
1. Client and server exchange data using TCP segments.
2. Sender sends segments; receiver acknowledges them.

3.Session Tear Down:

1. Either side initiates connection closure with FIN segment.
2. Receiver acknowledges FIN with ACK.
3. Receiver also sends FIN.
4. Sender acknowledges FIN with ACK.
5. Connection enters TIME_WAIT state.
6. After TIME_WAIT, connection is closed.

Overall Architecture

Network stack



IP input Heandler:

- MAC Protocol Parsing
- IP checksum verification
- IP Protocol Parsing
- Split Streams

ARP table

- Handling ARP requests and storing MAC-IP tuples

IP output handler

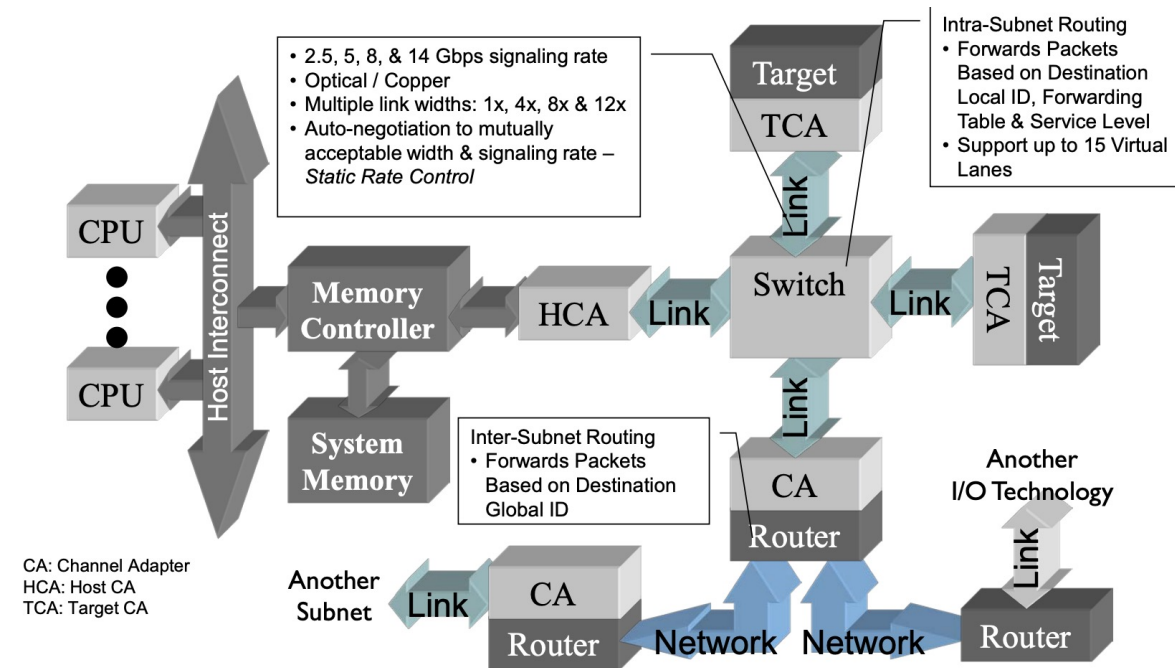
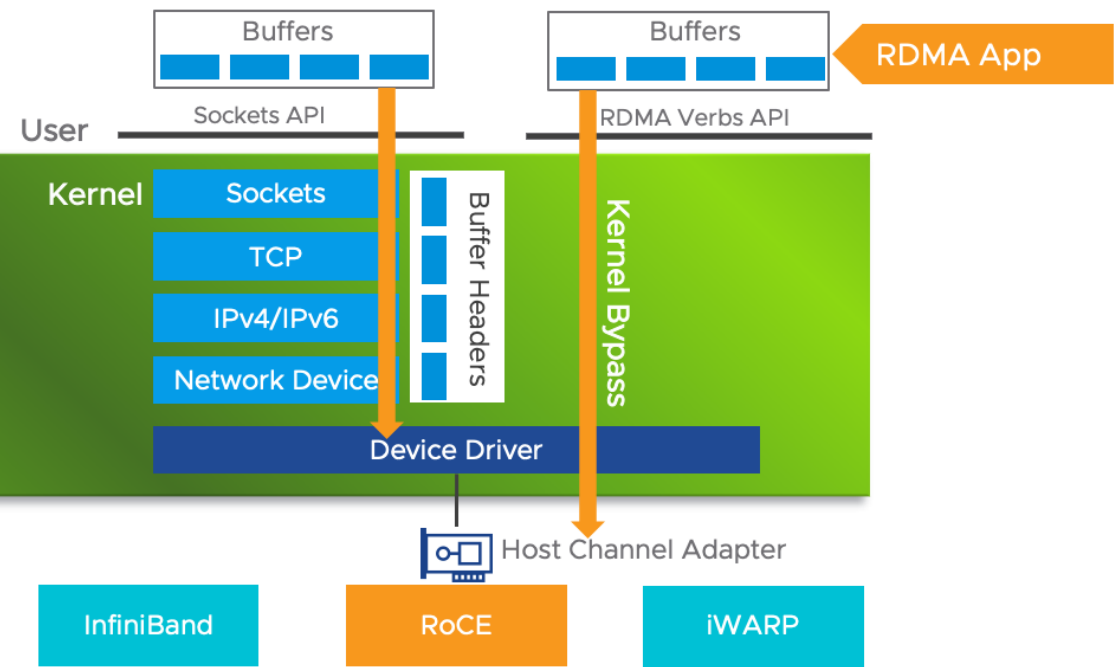
- Constructing MAC frame
- IP checksum computation
- MAC Lookup in ARP Table
- Merge Streams

ICMP

- Handling ARP requests and storing MAC-IP tuples

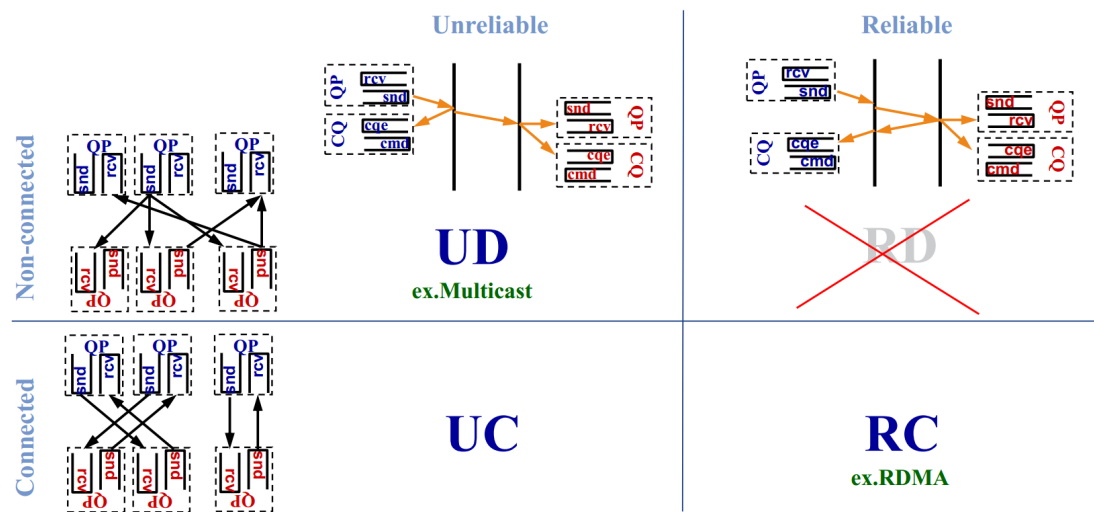
Overall Architecture

RDMA brief

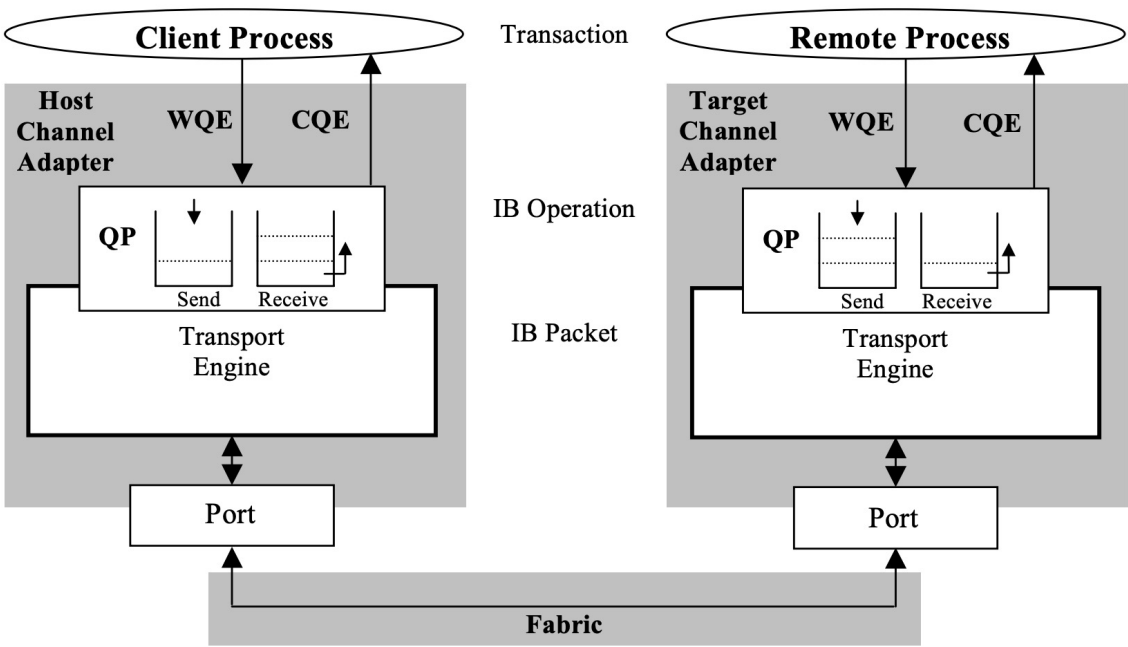


Overall Architecture

RDMA brief

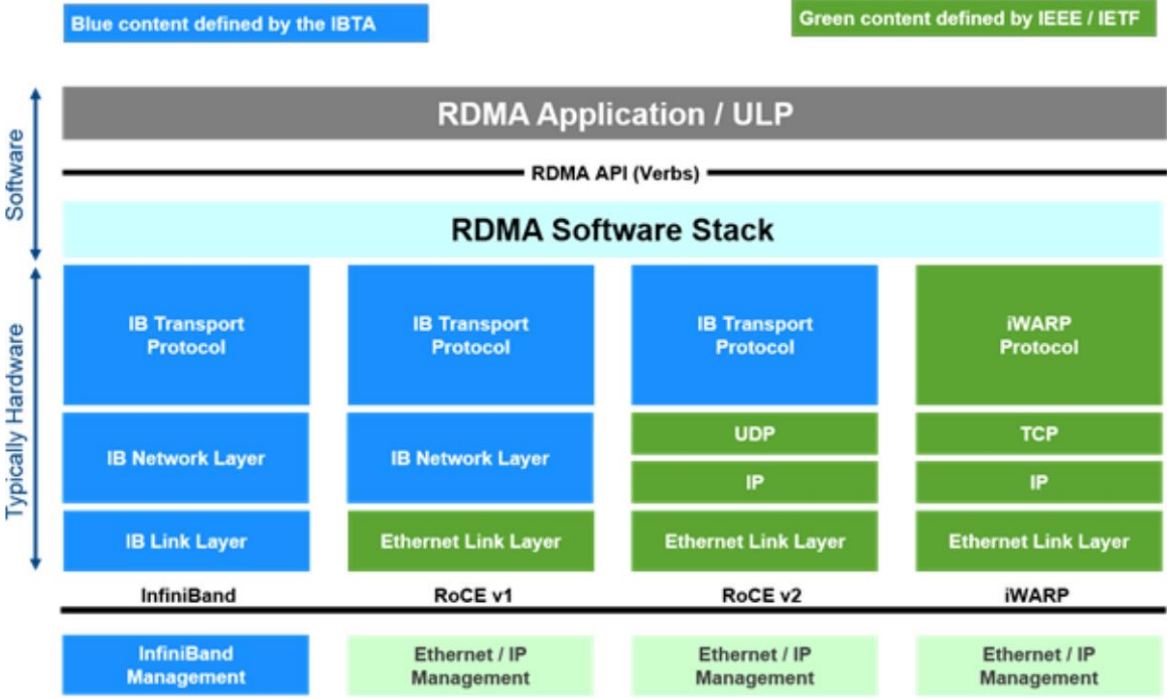
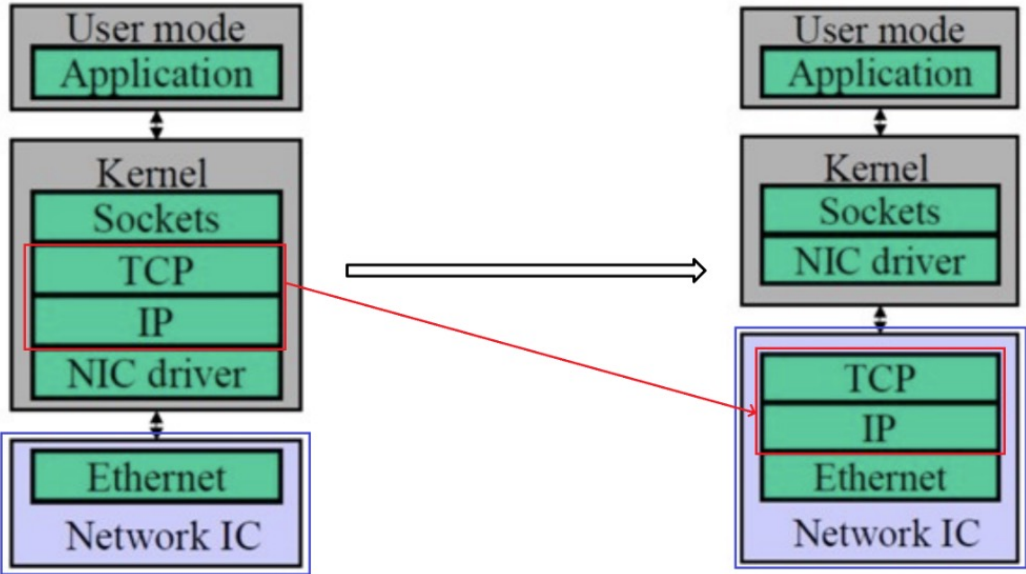


bits bytes	31-24		23-16				15-8		7-0		
0-3	OpCode		SE	M	Pad	TVer	Partition Key				
4-7	Reserved 8 (masked in ICRC)		Destination QP								
8-11	A	Reserved 7	PSN - Packet Sequence Number								



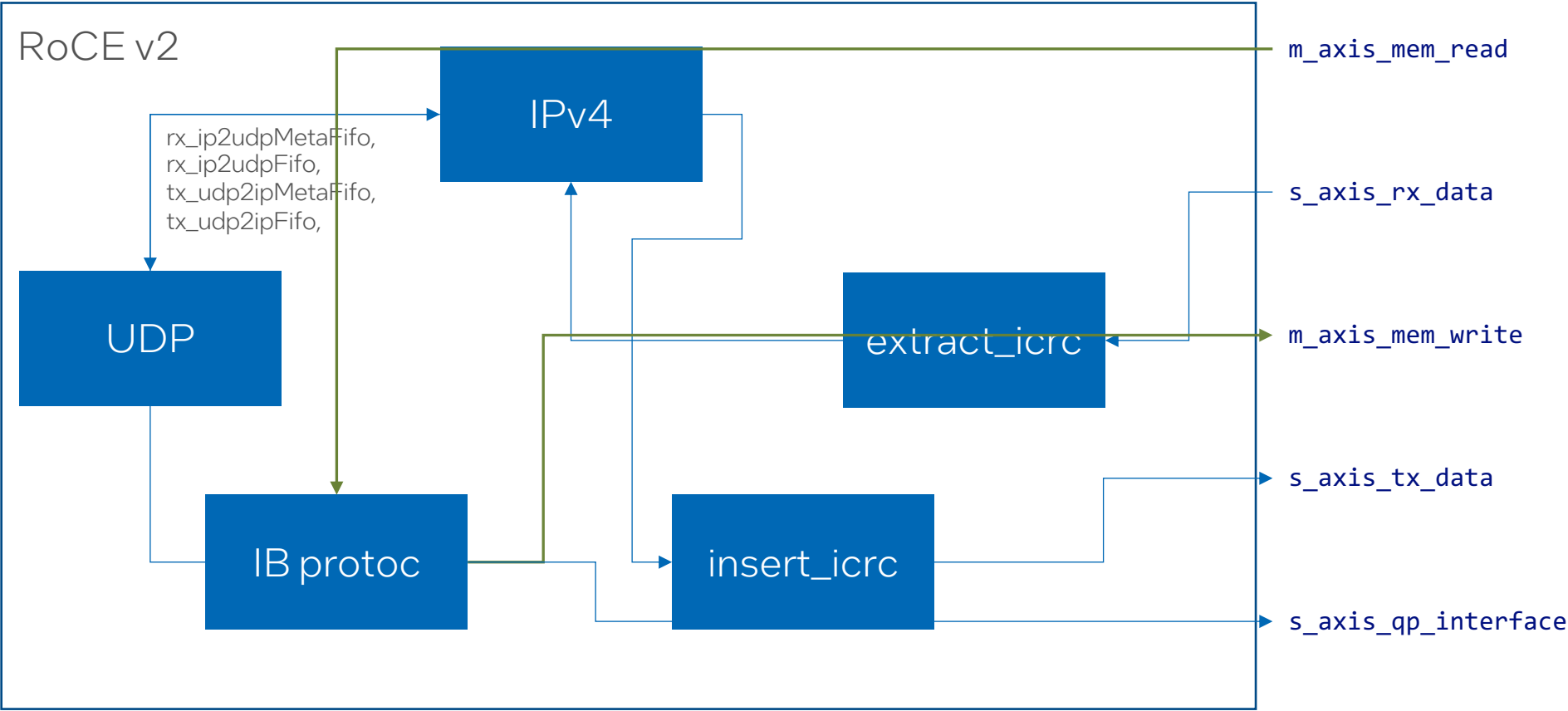
Overall Architecture

RDMA brief



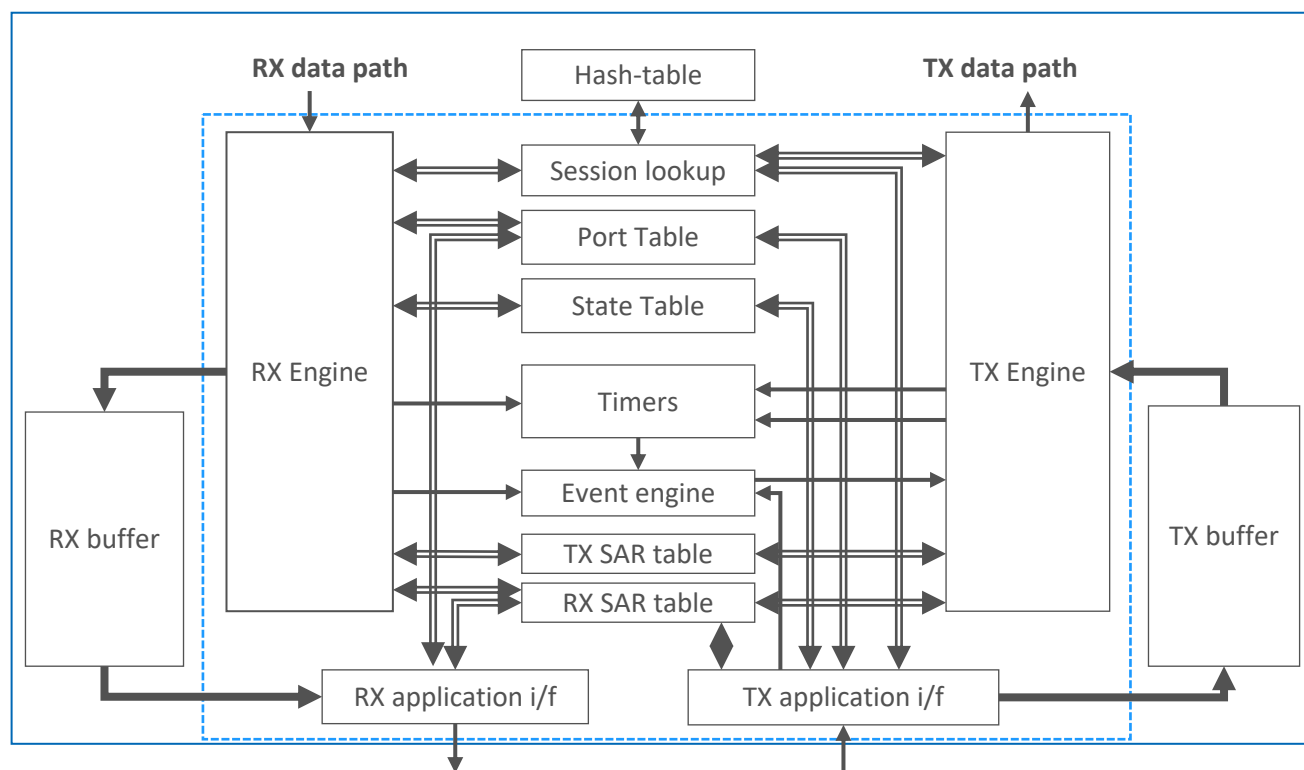
Overall Architecture

RoCE v2 Architecture



Overall Architecture

TOE architecture



Session lookup - Keeps Track of used and listening IP port

Event engine - Propagates Events through the system

Timers - for retransmission, probing, closing

State Table - Stores the state of each session: CLOSED, SYN-RECV, SYN-SENT, ESTABLISHED....

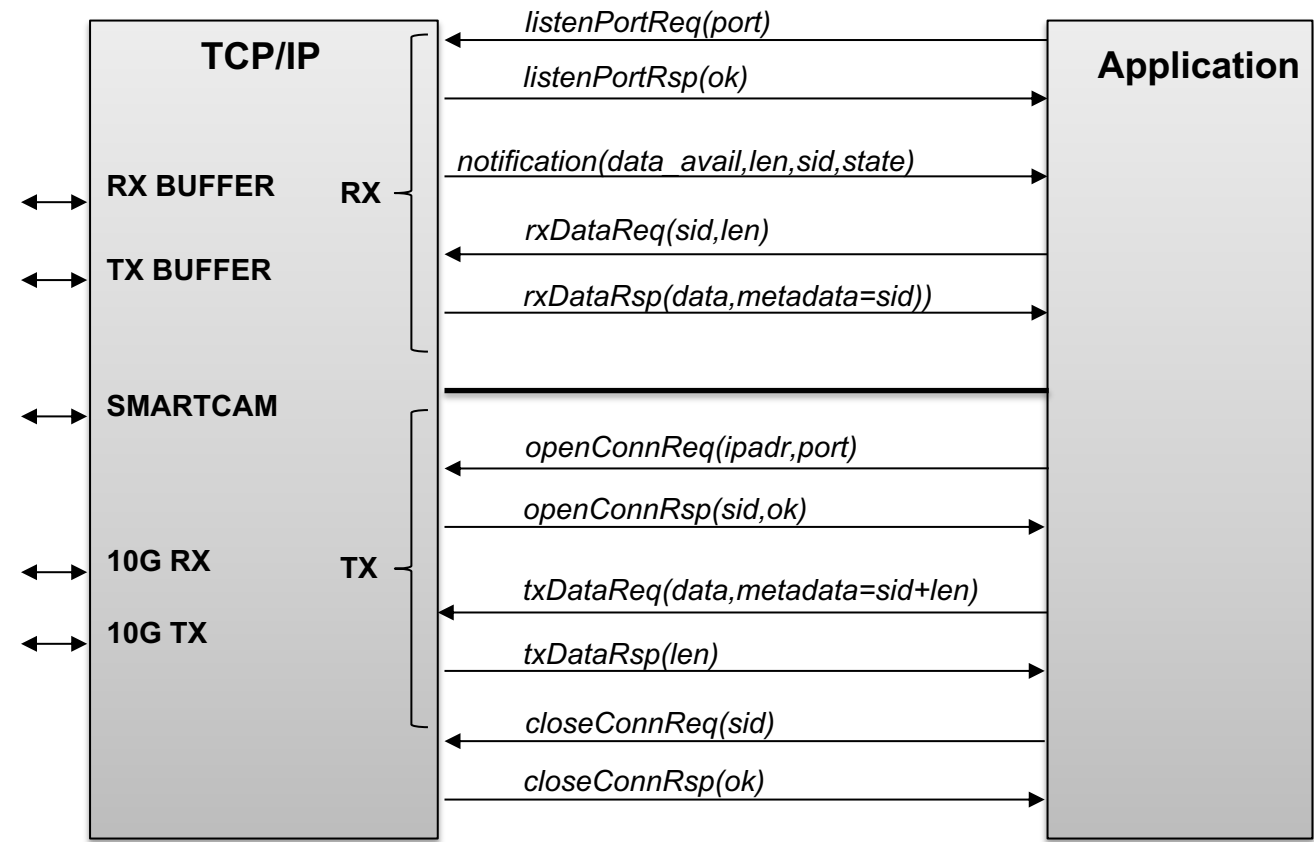
SAR Tables - storing the information of the TCP receive & send window

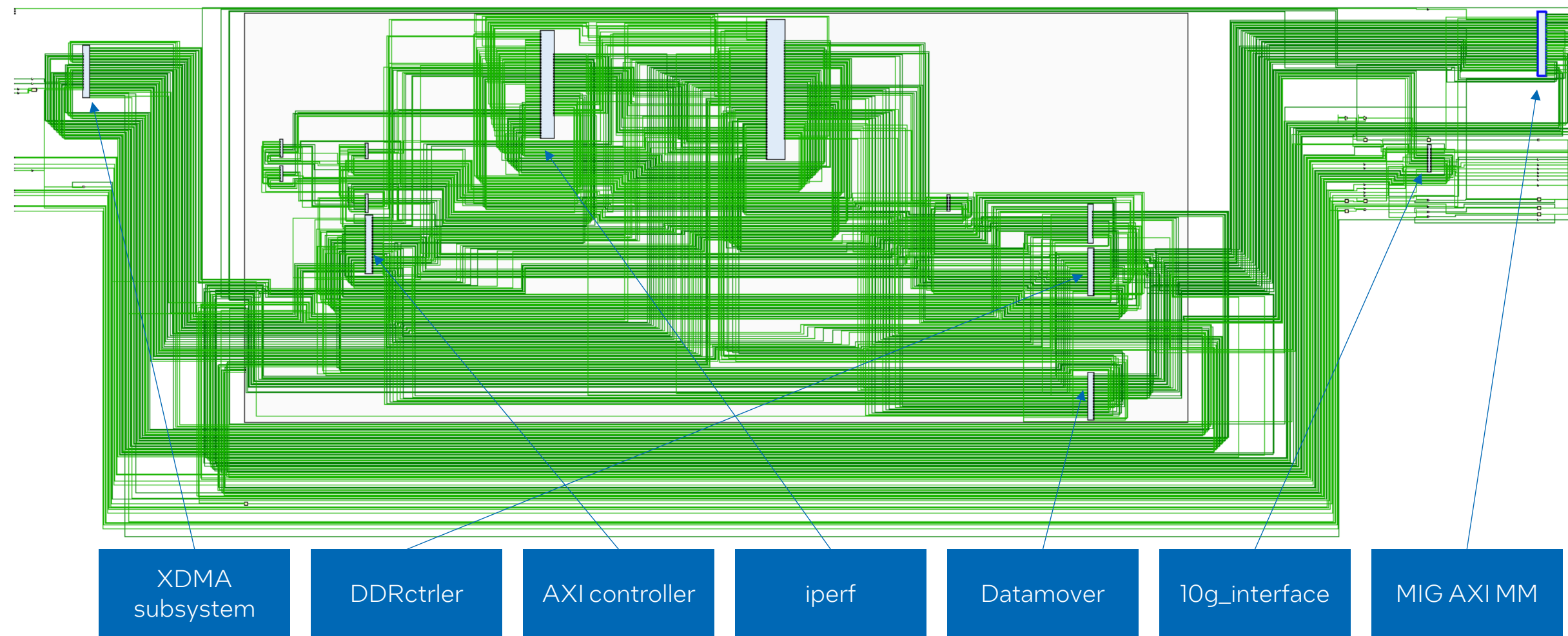
TX buffer - To support 10K sessions data is buffered in DDR memory (1.3Gbyte)

RX buffer – using streams.

Overall Architecture

TOE architecture





Overall Architecture

Performance

Latency / Line of Code

➤ Latency (1B payload):

- TCP Offload Engine:
 - Rx path: 1.1µs
 - Tx path: 0.8 µs
- UDP Offload Engine:
 - Rx path: 118 ns
 - Tx path: 92.4 ns

➤ Line of code:

HLS	IP Handler	ARP Server	MAC/IP Encoder	ICMP Server	TOE non-OOO	TOE OOO	UOE	Total
LoC	1592	462	370	534	5985	6441	736	10142

TCP/UDP Offload Resource Use

➤ All results are post-synthesis with Vivado 2014.2

	Network Interface	Memory Interface	TCP/IP Stack	Total	Available on V7690T
FF	5581	57637	28713	95873 11.1%	866400 100%
LUT	5321	43591	28706	81884 18.9%	433200 100%
BRAM	8	36	441	471 32.0%	1470 100%

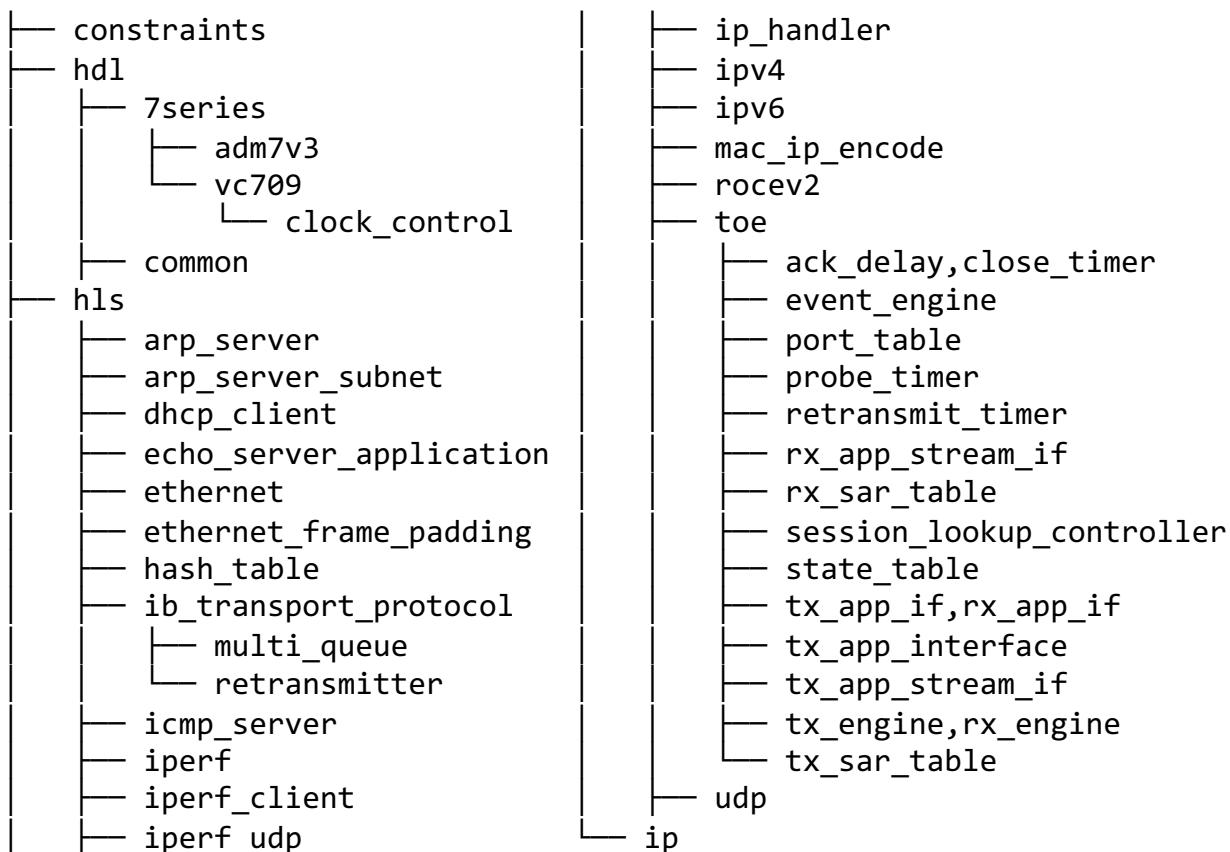
Breakdown (TCP/IP)

	IP Handler	ARP Server	ARP Smart CAM	MAC/IP Encode	ICMP Server	DHCP Client	Session Smart CAM	TOE non-OOO	TOE OOO	UOE
FF	1119	1251	1907	721	1568	505	2065	15504	17638	1939
LUT	1036	607	1653	566	1573	468	2112	14897	18698	1993
BRAM	3	6	42.5	9	12	0	57.5	250.5	304	16

HLS design

Subtitle Text Goes Here

components



HLS design

Conversion flow

Arbitrary precision integer.

1. In the search bar, type the following regular expression:
2. type `ap_uint<([>]+)>`
3. in the replace bar type:
4. type `ac_int<$1, false>`

ap_int assignment.

1. for RHS
 - a) in search bar: `([a-z . _]+)\(((^,)+)\s+([^\s])+\)`
 - b) in replace bar: `$1.template slc<($2) - ($3) + 1>($3);`
2. for LHS
 - a) in search bar: `([a-z . _]+)\(((^,)+)\s*([^\s])+\)\s+([^\s;]+)`
 - b) in replace bar: `$1.set_slc($3, $4);`

HLS design

Conversion flow

intel HLS

ihc::stream_in
ihc::stream_out

Function API	Description
T read()	Blocking read call to be used from within the component
T tryRead(bool &success)	Non-blocking read call to be used from within the component. The success bool is set to true if the read was valid. That is, the Avalon® - ST valid signal was high when the component tried to read from the stream. The emulation model of tryRead() is not cycle-accurate, so the behavior of tryRead() might differ between emulation and simulation.

vivado HLS

hls::stream

```
func1()  
{  
    while(!s.empty()) {  
        s.read();  
    }  
}  
func1()  
{  
    while(!tlast) {  
        s1.read();  
    }  
}
```

HLS design

```
template <class PipeName, class T, int depth>
class hls_stream
{
private:
    bool valid, wr_buffer_full;
    T latest_data, wr_buffer;
    ihc::pipe<PipeName, T, depth> member_pipe;

public:
    hls_stream() {
        valid = 0;
        wr_buffer_full = 0;
    }
    bool empty()
    {
        bool empty = !valid;
        return empty;
    }
    void pulse() {
        if (valid == 0)
            latest_data = member_pipe.read(valid);
    }
    T read()
    {
        valid = 0;
        return latest_data;
    }
};
```

```
void flush() {
    if (wr_buffer_full) {
        bool succ = 0;
        member_pipe.write(wr_buffer, succ);
        if (succ)
        {
            wr_buffer_full = 0;
        }
        else {
            wr_buffer_full = 1;
        }
    }
}

bool full() {
    return wr_buffer_full;
}

bool write(T data)
{
    if (!wr_buffer_full)
    {
        wr_buffer_full = true;
        wr_buffer = data;
        return true;
    }
    else {
        return false;
    }
}

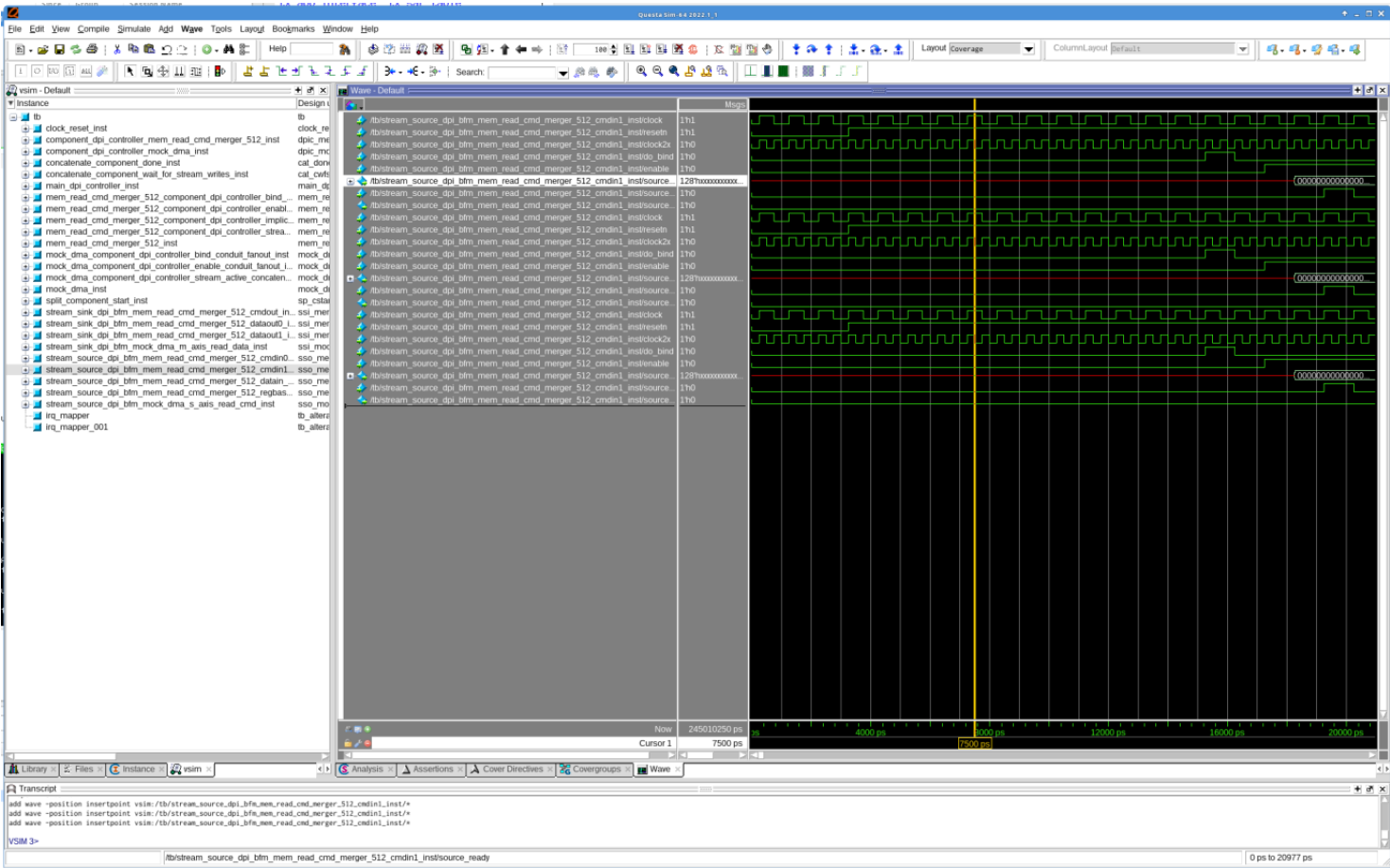
};
```

HLS tuning

dEBUG

HLS tuning

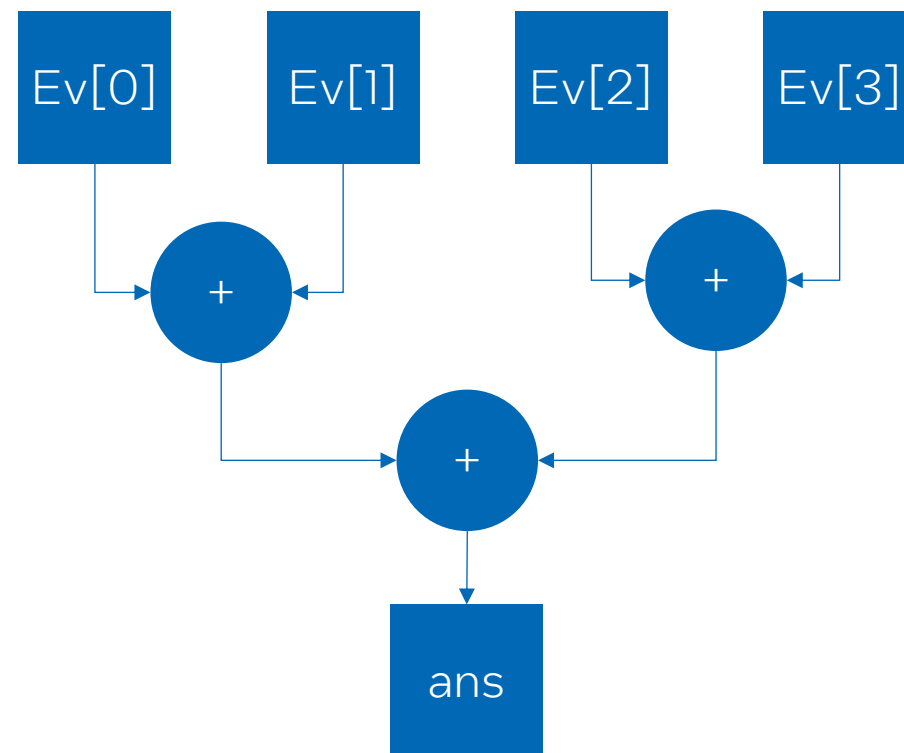
Debug - simulation



HLS tuning

Optimize ii

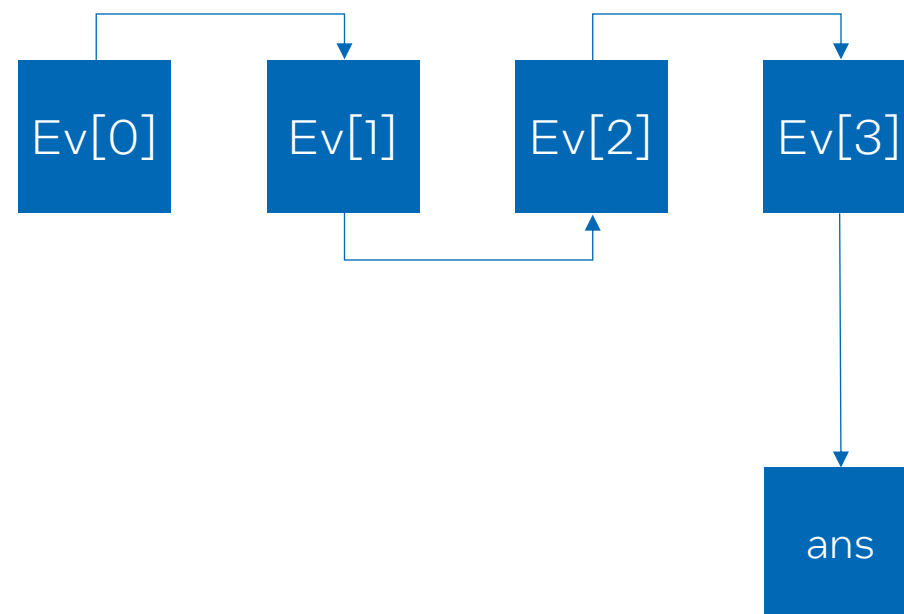
```
switch (state)
{
case RECEIVE:
    ev = input.tryRead(success);
    result = 0;
    #pragma unroll
    for(int i = 0; i < N; i++){
        result += ev[i];
    }
    state = PROCESS;
    break;
default:
    break;
}
```



HLS tuning

Optimize ii

```
switch (state)
{
case RECEIVE:
    ev = input.tryRead(success);
    result = 0;
    result = ev[0];
    state = PROCESS;
    counter = N - 1;
    break;
case PROCESS:
    #pragma unroll
    for(int i = 0; i < N-1; i++){
        ev[i] = ev[i+1];
    }
    result += ev[0];
    counter --;
    if(counter == 0) state = RECEIVE;
```



HLS tuning

example

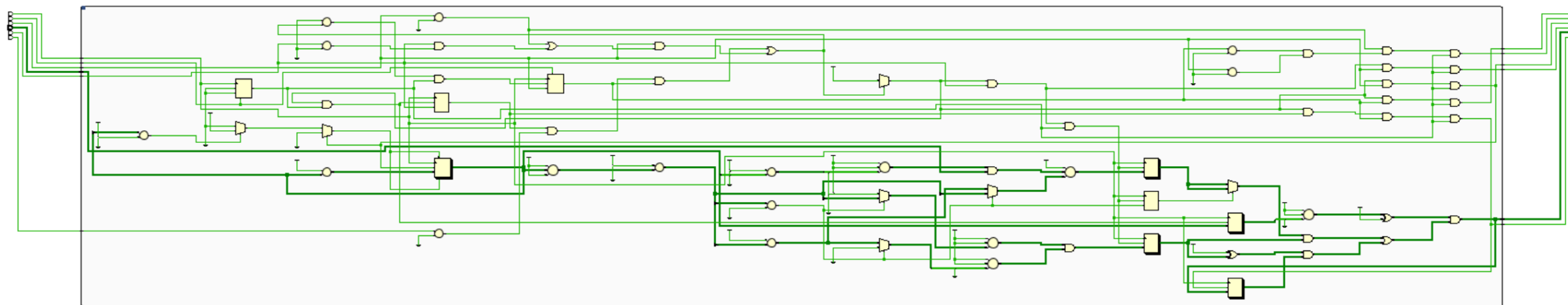
```
static ap_uint<32> ad_pointer = 0;
//static ap_uint<4> ad_readCounter = 0;
ap_uint<32> ev;
static int offset = 1;
if (!input.empty())
{
    ev = input.read();

    ad_pointer(WIDTH-1, WIDTH - (offset*8)) = ev(offset*8-1, 0);
    ad_pointer(WIDTH - (offset*8) -1, 0) = 0;

    output.write(ad_pointer);
}
if(offset < 4) offset++;
else offset = 1;
```


HLS tuning

example



HLS tuning

example

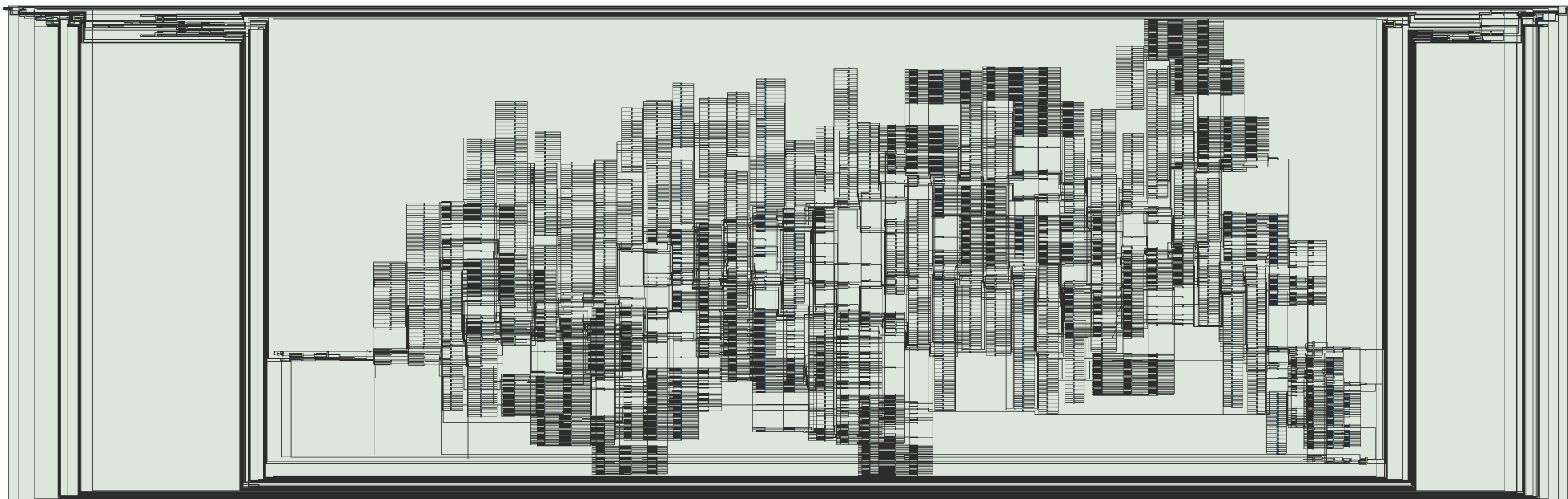
```
static ac_int<32, false> ad_pointer = 0;
//static ac_int<4, false> ad_readCounter = 0;
ac_int<32, false> ev;

static int offset = 1;
bool success;
ev = input.tryRead(success);
if (success)
{
    #pragma unroll
    #pragma ivdep
    for (int i = 0; i < WIDTH; i++)
    {
        if(i < offset*8) ad_pointer[i + WIDTH - (offset*8)] = ev[i];
        else ad_pointer[i] = 0;
    }

    output.write(ad_pointer);
}
if(offset < 4) offset++;
else offset = 1;
```

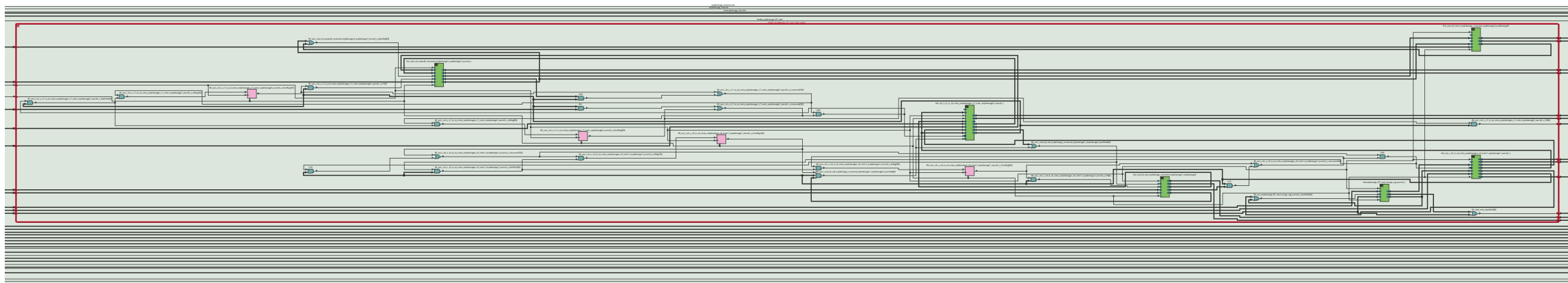
HLS tuning

example



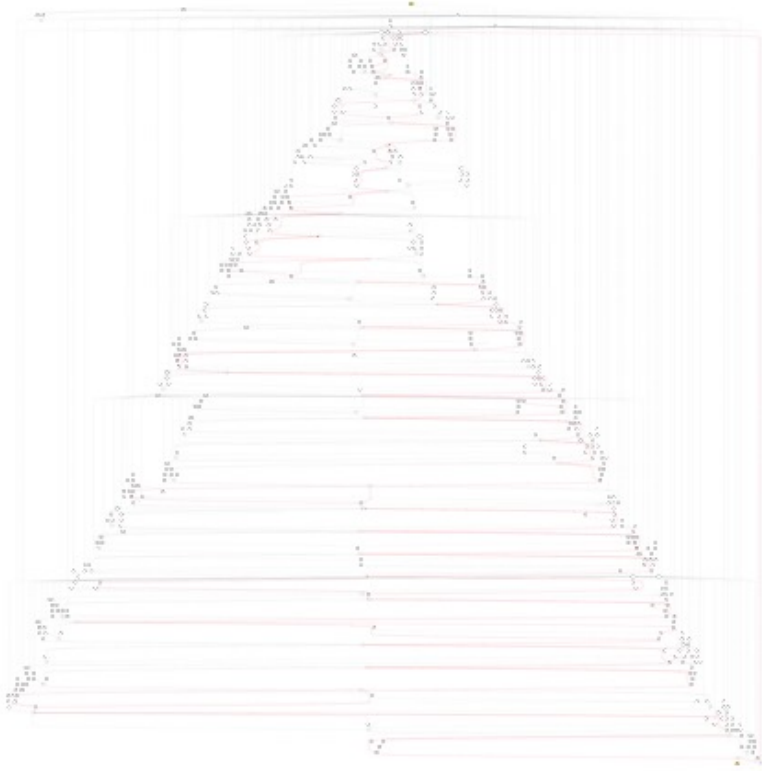
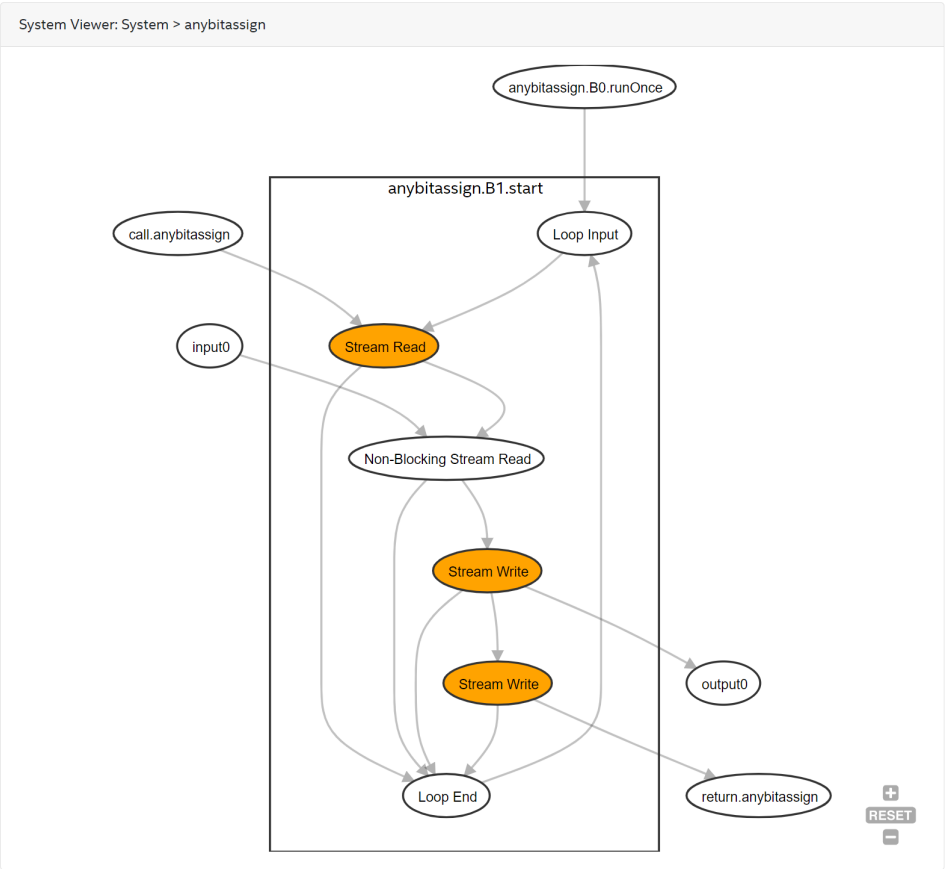
HLS tuning

example



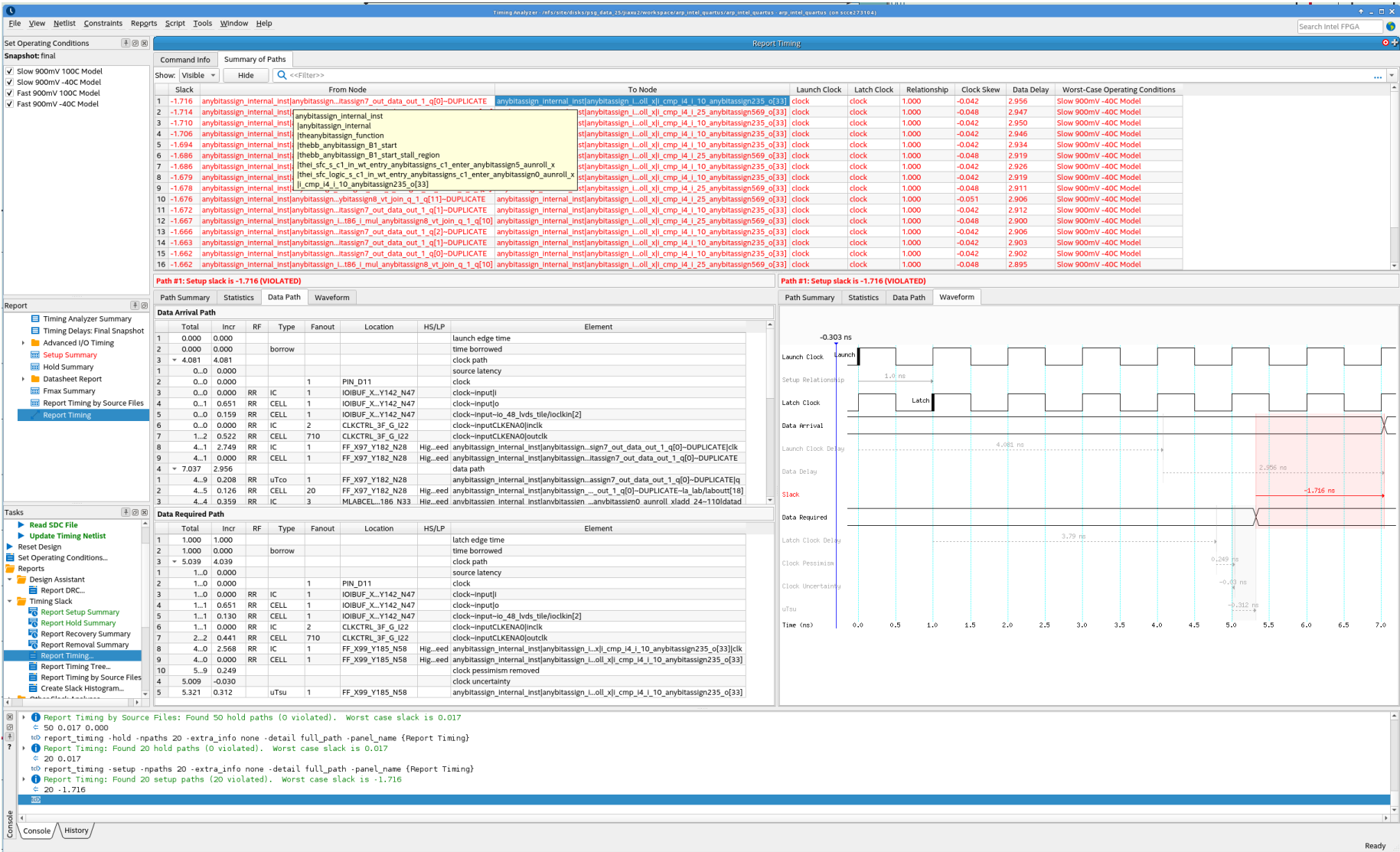
HLS tuning

example



HLS tuning

Frequency



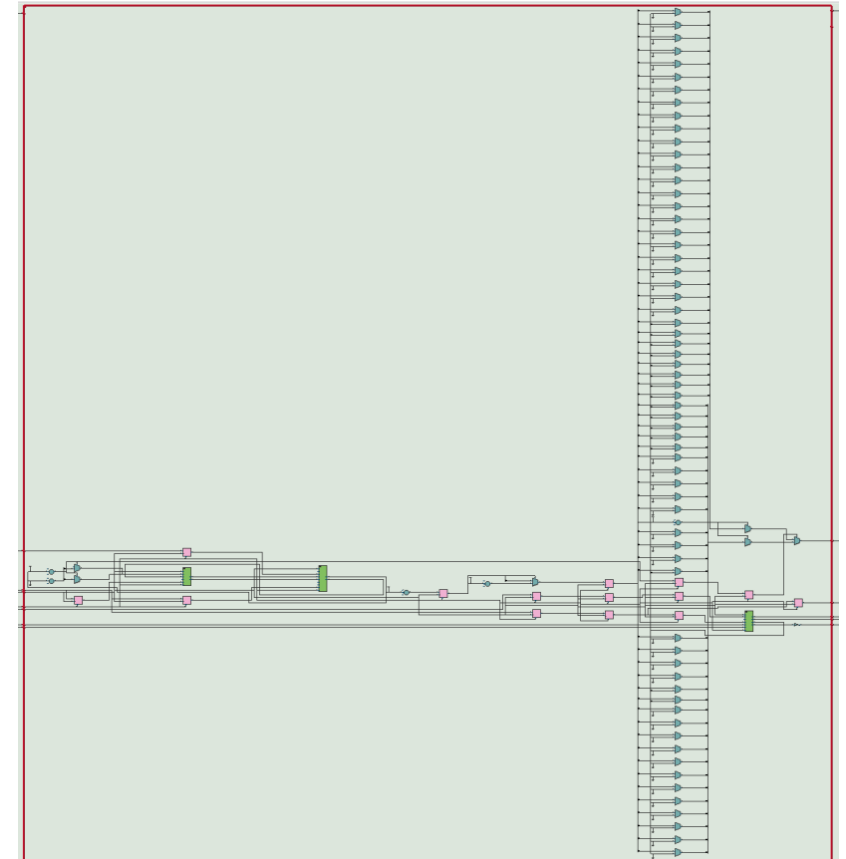
HLS tuning

example

```
static ac_int<32, false> ad_pointer = 0;
//static ac_int<4, false> ad_readCounter = 0;
ac_int<32, false> ev;

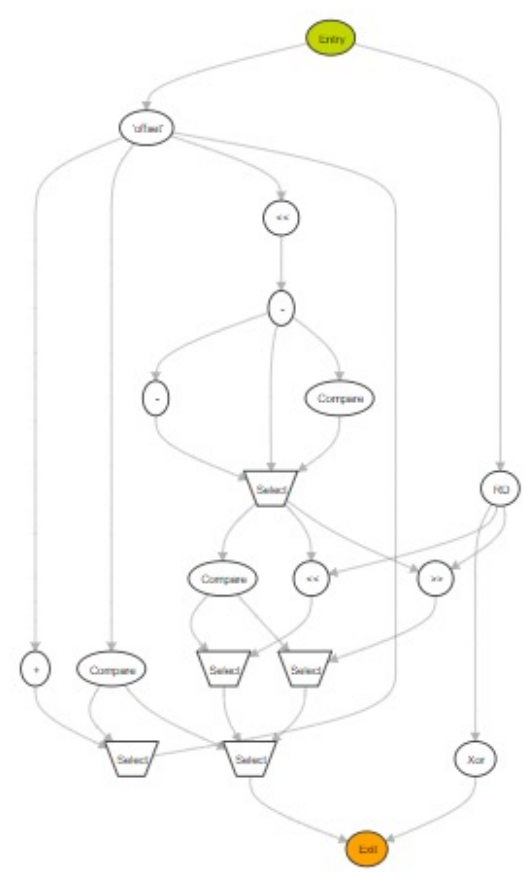
static int offset = 1;
bool success;
ev = input.tryRead(success);
if (success)
{
    ad_pointer = ac_int<32, false>(ev << (WIDTH - (offset*8)));

    output.write(ad_pointer);
}
if(offset < 4){
    offset++;
}else{
    offset = 1;
}
```



HLS tuning

example



Compilation Dashboard

Compilation Report - arp_intel_quartus

Table of Contents

- Parallel Compilation
- Flow Summary
- Flow Settings
- Flow Non-Default Global Settings
- Flow Elapsed Time
- Flow OS Summary
- Flow Log
- Synthesis
- Fitter
- Timing Analyzer
 - Summary
 - Timing Delays: Final Snapshot
 - Parallel Compilation
 - Clocks
 - Timing Closure Summary (BETA)
 - Fmax Summary

Fmax Summary

Show: Visible Hide <<Filter>>

	Fmax	Restricted Fmax	Clock Name	Note	Worst-Case Operating Conditions
1	397.77 MHz	397.77 MHz	clock		Slow 900mV -40C Model

The Intel logo is centered on a solid blue background. It features the word "intel" in a white, lowercase, sans-serif font. A small, light blue square is positioned above the letter "i". To the right of the word "intel" is a small white registered trademark symbol (®).

intel®