

Development of 10Gb Ethernet applications in FPGAs FPGAworld Conference 2014- Copenhagen, Denmark September 2014 Steinn Gustafsson





#### The Objective: Move large amounts of data quickly and reliably from PC to FPGA



Candidate protocols

Ethernet

- DVI, HD-SDI
   High data rates but not ultra-reliable, not error free
- USB, SATA Reliable but not very high throughput, short range
  - PCI
     Reliable, high throughput, short range
    - (TCP/UDP) High throughput, long range





# **10 Gigabit Ethernet**





#### **CHEVIN** TECHNOLOGY

# The PHY Layer 1

•IEEE802.3ae-2002

•XGMII –10G Media Independent Interface, 32bit @312.5MHz.

•64B66B Encoding & Scrambling - EMI reduction

- •Gearbox 10Gbps <-> 10.3125Gbps
- •Clock Rate adaptation handle +/-100ppm difference

•PMA – Parallel<->Serial, CDR, Gigabit transceivers







# The MAC Layer 2

•XGMII –10G Media Independent Interface, use any PHY, 32bit @312.5MHz
•XAUI/XGXS "XGMII extension Sub-layer" between MAC /PHY, e.g. an external IC, backplane.
•XAUI rate 4 pairs @ 3.125GHz each way. 8b/10b
•Application – Streaming Interface (user specified, e.g. Avalon,AMBA,other)







#### **Store&Forward vs Cut-through**

•Store-and Forward. No bad frames leave the MAC. Variable and longer latency •Cut-through. Fixed and shortest latency. Possible "clean-up" required at App





# Transport Layer 4 Protocols

- TCP connection oriented
  - "Stream of bytes" between applications over sockets
  - Guaranteed delivery
  - Correct Order
  - Server and Client roles
  - Complex protocol, 1GHz CPU resources for each 1Gbps TCP traffic
  - Point-to-point only

#### UDP connection-less

- "message oriented" frames between applications over sockets
- Uncertain delivery "Send and forget"
- No re-ordering
- UDP message remains intact
- Simple protocol, no sequence number, no ACK
- Broadcast/Multicast/Unicast





### **The UDT Protocol Layer 5**

•UDT4 (UDP based Transfer Protocol) Yunhong Gu, University of Illinois at Chicago 2007
•Light weight – UDP with Sequence numbers
•Sliding Receive Window
•Lost packets are sent individually

Identical nodes maintain Loss listCongestion Control



TECHNOL





#### **UDT Server in FPGA**

#### **Offload Engine Tasks**

- Separate UDT Payload from UDT header process separately CPU / RTL
- Move UDT Header to Buffer Microblaze CPU processes UDT protocol Header
- Move UDT Payload to Buffer RTL processes UDT protocol Payload
- Microblaze CPU sends response ACK, maintains & sends loss list, etc





#### **Verification using Simulation**

- Simulation Modelsim
- Vector tools
  - pcap2sim("xgmac\_10g\_test.pcap",[1,(3..25)],"vector.bin" )
  - sim2pcap("sim\_result.bin" ,"sim\_result.pcap")
- Unit Tests
  - MAC/IP Address
  - PING, ARP
  - Filtering, UDT server





#### **Verification on Hardware**

- Traffic Analysis Wireshark
- Send something to the FPGA
  - > send.exe filename 192.168.1.102 9000
- Unit tests verify data is sent correctly to FPGA
  - Rate >9.5Gbit/s, error-free
  - Errors, bit errors, dropped packets
  - Flow Control
  - "Soak test", check DDR3 memory content

🔼 xcma	ac_10g_test.pcap	[Wireshark 1.6.5 (SVN Rev 40429 from /trunk-1.6)]		-	A TANK ( ) & H TA A March of State March ( ) State	
<u>F</u> ile <u>E</u>	dit <u>V</u> iew <u>G</u> o	Capture Analyze Statistics Telephony Tools Inter	nals <u>H</u> elp			
		🗀 🖬 🗶 😂 占   🔍 🗢 🔶 春 👱	E 🗐 🔍 Q 🔍 🗹	🗃 🖂 🐔 🔅		
Filter:		•	Expression Clear Apply			
No.	Time	Source	Destination	Protocol	Length Info	
	1 0.000000	Qlogic_05:72:58	Broadcast	ARP	42 who has 192.168.10.2? Tell 192.168.10.1	
	2 0.000068	aa:bb:cc:dd:ee:ff	Qlogic_05:72:58	ARP	60 192.168.10.2 is at aa:bb:cc:dd:ee:ff	
	3 0.000076	192.168.10.1	192.168.10.2	UDT	106 Control, Handshake, Sock: 0	
	4 0.000131	192.168.10.2	192.168.10.1	UDT	106 Control, Handshake, Sock: 1fab99bb	
	5 0.000183	192.168.10.1	192.168.10.2	UDT	106 Control, Handshake, Sock: 0	
	6 0.000230	192.168.10.2	192.168.10.1	UDT	106 Control, Handshake, Sock: 1fab99bb	
	7 0.001374	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: 1, Seq: 712b6b12, Sock: 11223344	
	8 0.001386	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: 2, Seq: 712b6b13, Sock: 11223344	
	9 0.001394	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: 3, Seq: 712b6b14, Sock: 11223344	
	10 0.001402	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: 4, Seq: 712b6b15, Sock: 11223344	
	11 0.001409	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: 5, Seq: 712b6b16, Sock: 11223344	
	12 0.001416	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: 6, Seq: 712b6b17, Sock: 11223344	
	13 0.001423	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: 7, Seq: 712b6b18, Sock: 11223344	
	14 0.001430	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: 8, Seq: 712b6b19, Sock: 11223344	
	15 0.001436	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: 9, Seq: 712b6b1a, Sock: 11223344	
	16 0.001443	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: a, Seq: 712b6b1b, Sock: 11223344	
	17 0.001443	192.168.10.2	192.168.10.1	UDT	82 Control, ACK, Sock: 1fab99bb	
	18 0.001449	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: b, Seq: 712b6b1c, Sock: 11223344	
	19 0.001459	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: c, Seq: 712b6b1d, Sock: 11223344	
	20 0.001468	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: d, Seq: 712b6b1e, Sock: 11223344	
	21 0.001478	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: e, Seq: 712b6b1+, Sock: 11223344	
	22 0.001486	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: †, Seq: 712b6b20, Sock: 11223344	
	23 0.001495	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: 10, Seq: /12b6b21, Sock: 11223344	
	24 0.001496	192.168.10.2	192.168.10.1	UDT	82 Control, ACK, Sock: 1tab99bb	
	25 0.001514	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: 11, Seq: /12b6b22, Sock: 11223344	
	26 0.001533	192.168.10.1	192.168.10.2	UDT	8250 Data, Msg: 12, Seq: /1206b23, Sock: 11223344	
4						•



#### **Designing for 10G in FPGA**

- PHY XGPHY new possibilities in FPGA with SERDES > 10G (StratixV, Virtex7). XAUI option for slower SERDES FPGAs.
- MAC XGMAC easy to integrate, lean on resources, low-latency
- IP/TCP XGTCP easy to integrate, lean, optimised for few connections
- IP/UDP UDT Server easy to integrate, optimised for high bandwidth & distance





#### **Thank You**



www.chevintechnology.com

