400GbE FEC Discussion

Tongtong Wang, Xinyuan Wang

IEEE 802.3 400 GbE Task Force

Contributor

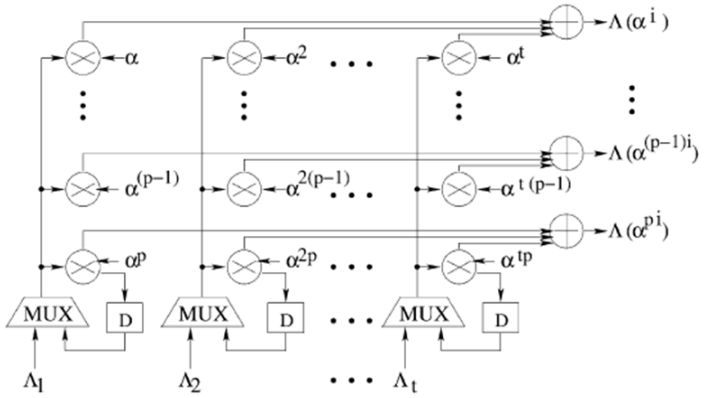
- Zhongfeng Wang, Broadcom

Hot Discussion on 400GbE RS FEC

- Performance
  - 4x100G FEC with FOM has best performance against burst error with lane identity limitation
  - 1x400G FEC dilute errors over multiple lanes but can't help relax system requirement
  - 1x400G nonFOM
  - 4x100G nonFOM
- Implementation friendly
  - 1x400G concerns
  - 2x200G
  - 4x100G
- Cost (area/latency) Comparison
- Enable potential features
  - Breakout

Concern 1: 4x Parallelism in 1x400G

- Every stage of calculation (Syndrome/Chien/Forney/Matrix Multiplier in Encoder) in RS FEC needs 4x parallelism, which means not only wider data width, but also longer critical path with larger fan out.



**Chien search engine, parallelism = p*

- For ASIC@~664MHz: parallelism is 64
- For FPGA@~332MHz: parallelism is 128

  Very Challenging for implementation in FPGA.

  Timing /PAR congestion problems?

- Timing converging is harder for encoder, that has critical path with more Galois Field Multiplier.

***Refer to <Small area parallel Chien search architectures for long BCH codes>,** Chen, Yanni; Parhi, K.K. Very Large Scale Integration (VLSI) Systems, IEEE Transactions on, May 2004.*

Concern 2: Half Cycle Problem in 1x400G

- 1x400G Option 1: **Use 680b bus width**
- KP4 RS FEC algorithm working on 680-b parallelism in ASIC (1360-b parallelism in FPGA), process each codeword in 8 (/4) cycles. Two sets of gearboxes are needed

- Extra Latency cost : > 2 + 6 cycle
- Area cost :
  - in ASIC 68P engine + 2x 68P gearbox
  - When 16 x25G, 42.5b/lane costs extra latency and design complexity
  - in FPGA 136P engine + 2x 136P gearbox

- More CDC(clock domain crossing) problem involved.

Why gearboxes are extra in 680b scheme?

- There is always a small frequency difference (typically ~+/-100ppm) between reference clock sources, even if they are nominally the same frequency. And this small difference will cause FIFO underrun or overflow condition.

- So in high speed transceivers, to allow clock correction with insertion/remove pad bits, usually use only one gearbox after PCS layer where idle frames are natural and easy to insert/remove.

- To use gearbox before FEC decoder as in *wilkie_01_0615_logic* is possible, but it is not generically needed and may cause other complexities.
  - *Insert/remove pad data in between codewords? Or pause the PCS function for the off duty cycles?*

- The gearbox after FEC decoder, converting from 680b back to 640b, can not always be combined together with transcoding module, in the cases when FEC/PCS is not in host ASIC as in *dambrosia_3bs_02b_0115 p8,p9*

Concern 2: Half Cycle Problem in 1x400G

- 1x400G Option 2: **Use 640b bus width**

- KP4 FEC algorithm working on 4x160b parallelism, process each codeword in 8.5 cycles@664MHz
  - At least one additional half engine is needed on the split cycle for Syndrome, Chien/Forney, and Encoder
  - Using 1280b at 332Mhz clock, at least one additional half engine is needed in the split cycle for Syndrome, Chien/Forney, and Encoder.

- Extra Latency cost : 3~4 cycles (x1.6ns, or x3.2ns)
- Area cost :
  - in ASIC 64P engine + 1x32p engine
  - in FPGA 128P engine + 1x64p engine

2x200G Implementation

- 2x Parallelism , less stress in timing convergence
- Use 2x320b data width over 16 lanes
  - No half cycle problem
  - Less CDC(clock domain crossing) problem
  - Every FEC engine distribute to 8 lanes.
  - 4x100G Breakout need time sharing or extra HW

4x100G Implementation

- Reuse 100G IP, shorter time to market

- Use 4x160b data width over 16 lanes
  - No half cycle problem
  - Less CDC(clock domain crossing) problem
  - Every FEC engine distribute to 4 lanes, same AM size as in .bj

Area/Latency Cost Comparison

Latency		
	Decoder	Encoder
4x100G	110ns	3ns
2x200G*	84ns	3ns
1x400G	72ns + 3*1.6ns	3ns+ 2*1.6ns

Latency difference is not significant with comparing to fiber latency @5ns/m

Area Cost(Decoder)				
	Synd	KES	CS	Forney
4x100G	4	4	4	4
2x200G	4	4	2	4
1x400G* (640b scheme)	4*(1.5)	4	1*(1.5)	4

Notes: (1) Assume half engine cost 50% extra area;
(2) encoder and control circuitry are not listed

Conclusion:

- From implementation friendly perspective, prefer to use 4x100 or 2x200 than 1x400G, particularly when 4x100G breakout is considered, where significant extra hardware/latency is involved for 1x400G solution.

Thank you