



# 100Gb/s Single-lane SERDES Discussion

Phil Sun, Credo Semiconductor

January, 2018

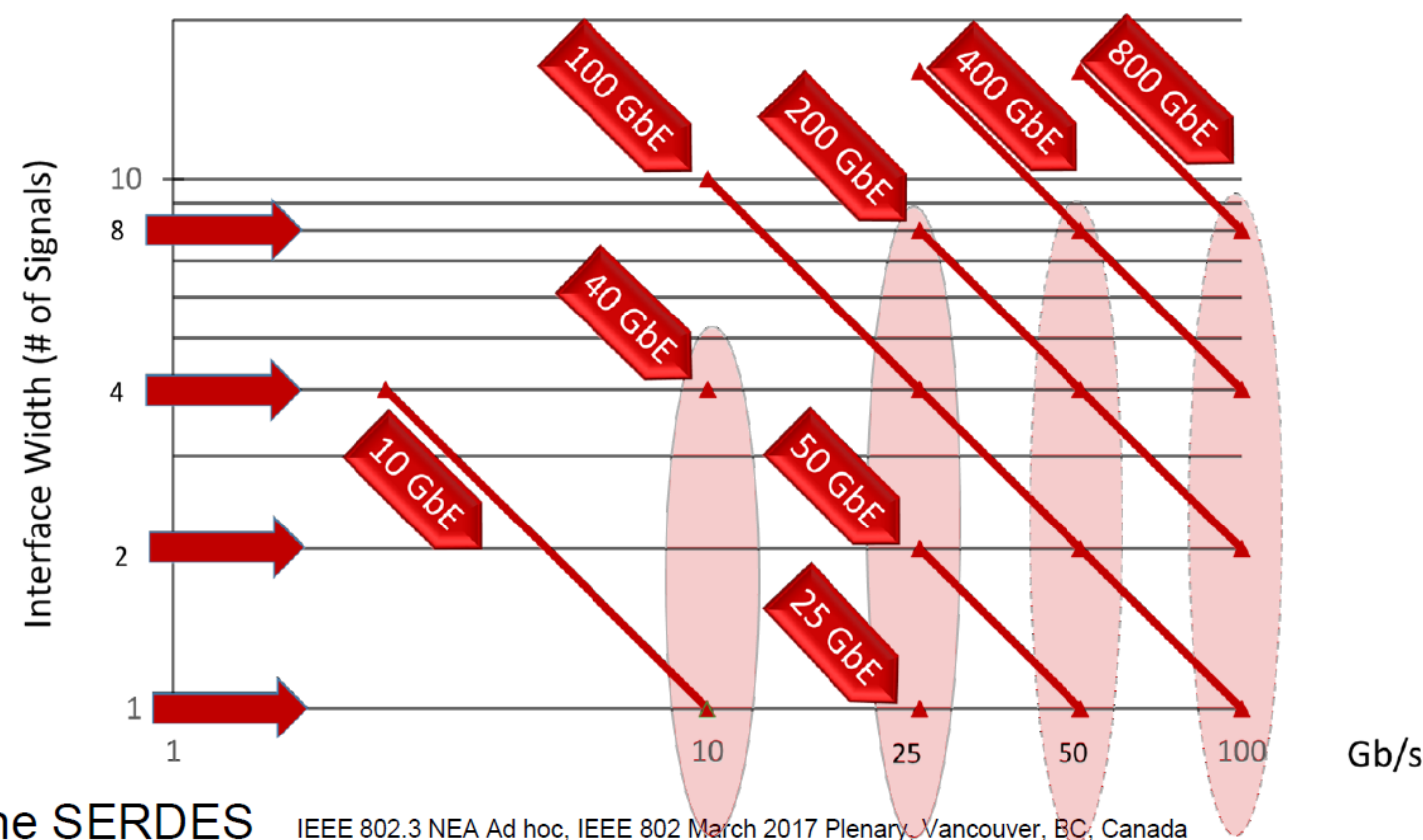
# Introduction

This contribution tries to share thoughts on 100Gb/s single-lane SERDES development and bring discussions on these topics:

- 100Gb/s SERDES Opportunities and Challenges
- Modulation choices: PAM4 v.s. PAM8
- BER Requirement and FEC
- Lower-power Architecture for 100Gb/s Long Reach SERDES
- TX FIR Training
  - Real-time tuning
  - TX Training time

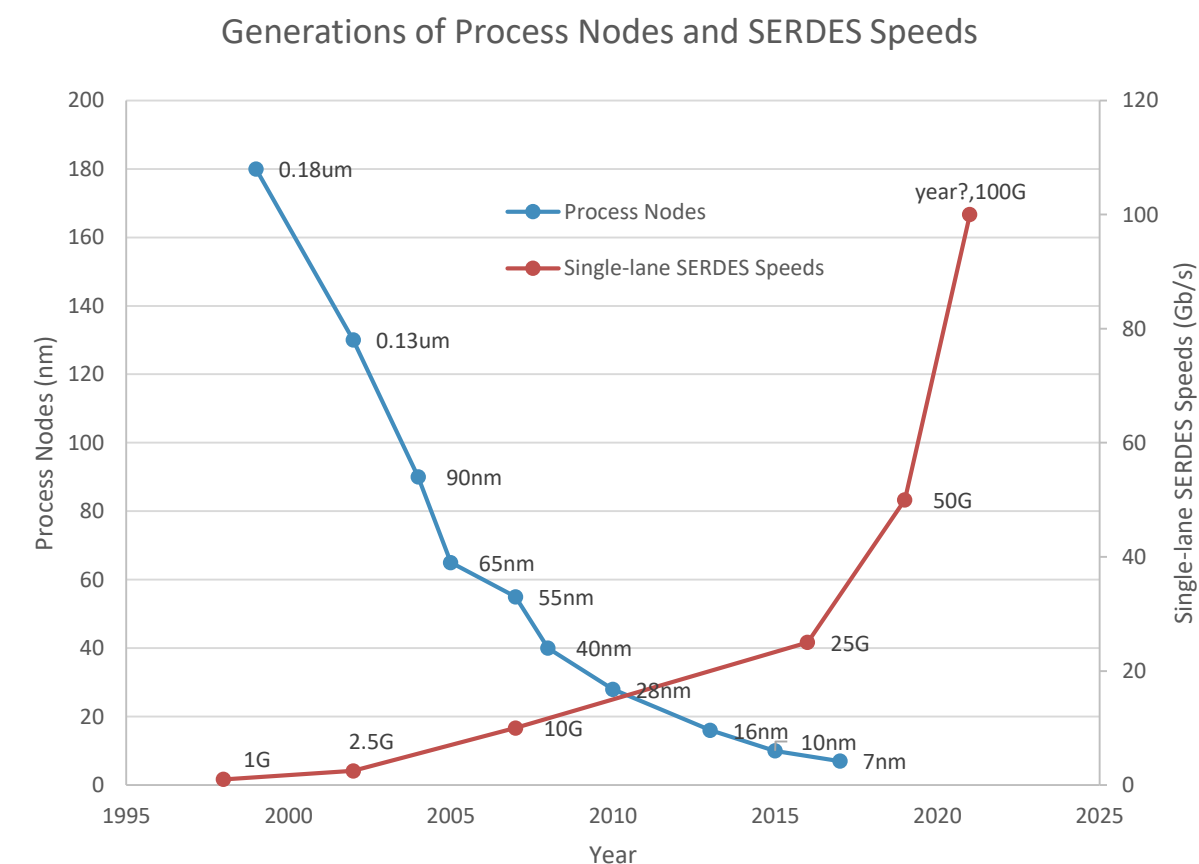
# 100Gb/s SERDES Opportunities and Challenges

- Higher speed SERDES is desired for higher throughput interconnect. On the other hand, it requires faster and more complexed circuits.
- SERDES design takes advantage of faster process nodes to solve design challenges and meet power constraints.
  - 100Gb/s short reach SERDES has been demoed on 28nm. Lower power may be achieved on 16nm and 7nm.
  - 100Gb/s long reach has higher complexity.



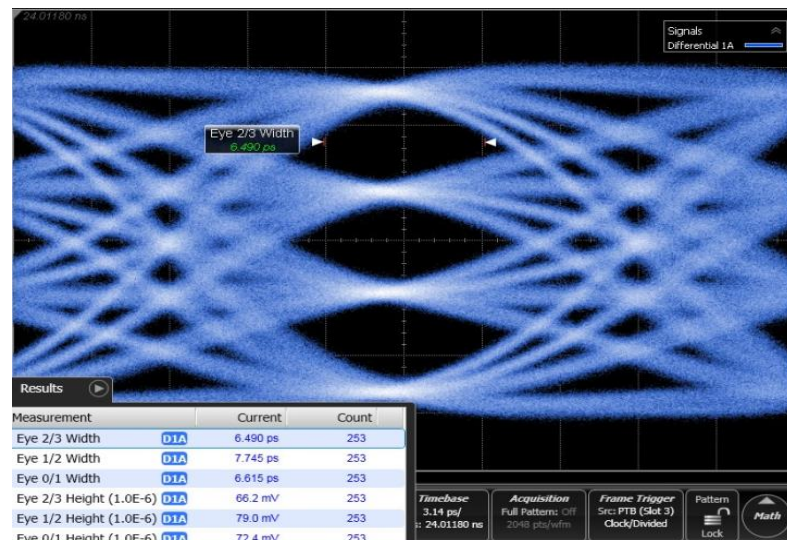
Follow the SERDES IEEE 802.3 NEA Ad hoc, IEEE 802 March 2017 Plenary, Vancouver, BC, Canada

[Goergen\_nea\_01a\_0317].

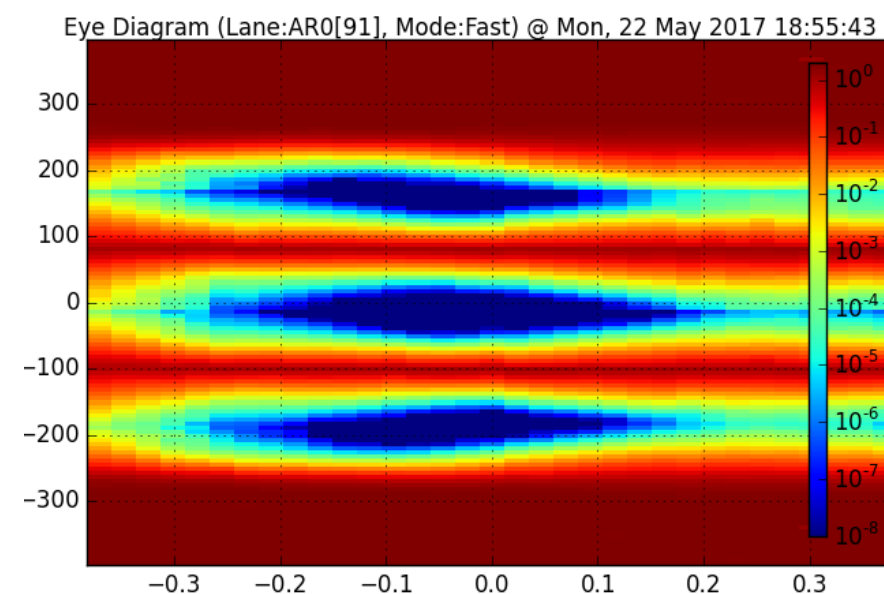


# 100G Short Reach Design Results

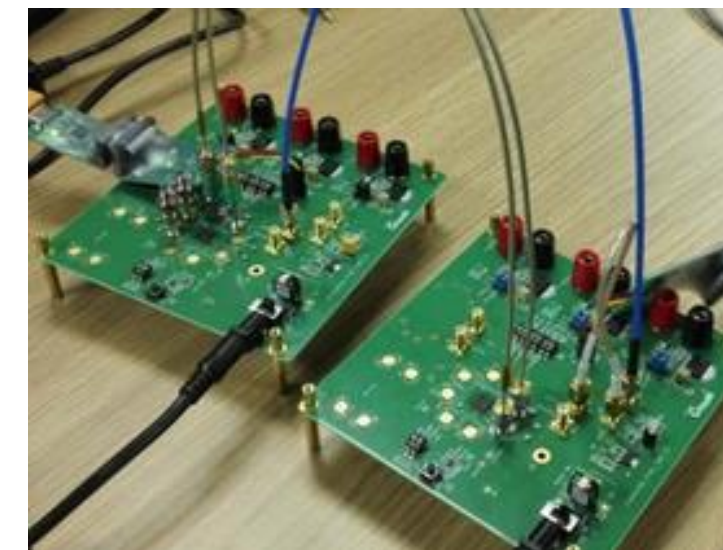
- A 100Gb/s PAM4 SERDES for short reach has been developed and demoed.
- With 28nm process node, TX eye is clean. Multiple tap TX FIR has been applied for TX eye measurement.



TX Eye Diagram



Eye Monitor



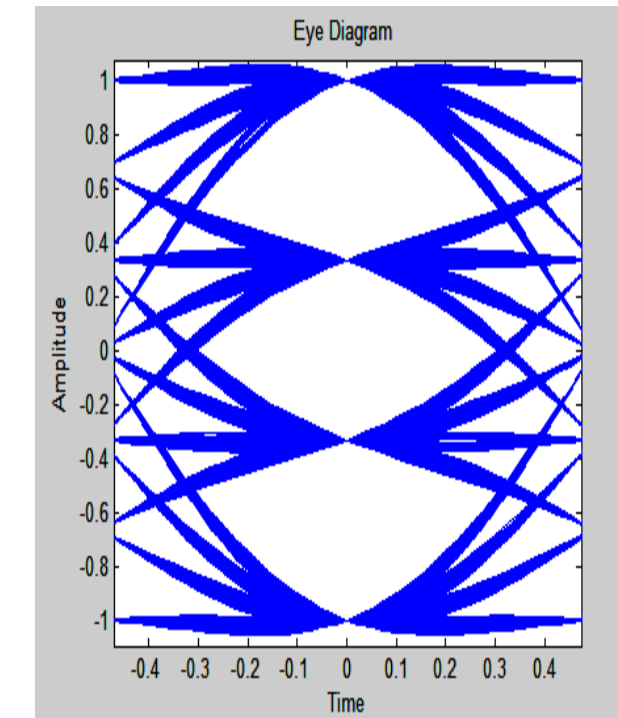
Test Setup



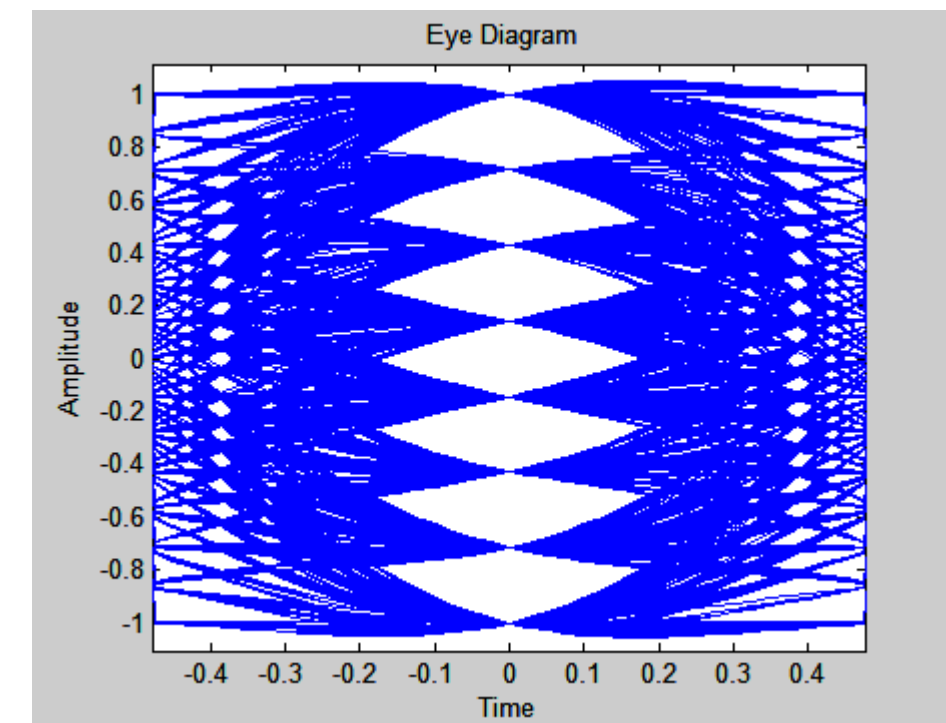
# Modulation Choices: PAM4 vs. PAM8

Considering high clock rate of SERDES and power/latency constraints, some hardware costly equalization and FEC schemes are unlikely to be used for 100GE. This contribution compares two modulation schemes assuming SERDES RX DFE (at least 1 tap) may be used and FEC power/latency should not be dramatically increased.

- From PAM4 to PAM8, bandwidth reduction is 1/3.
  - Less than 1/2 bandwidth reduction from NRZ to PAM4.
- PAM8 eye height is -7.4dB lower. Therefore it is more sensitive to residual ISI and circuit distortion.
- For PAM8, DFE error propagation rate is higher (7/8 v.s. 3/4), and each FEC symbol covers less (2/3) PAM8 symbols. Burst error penalty is worse for FEC (e.g. Reed Solomon FEC).
- For the FEC schemes shown later, DER requirement for PAM8 and PAM4 is 2.6E-8 and 3.8E-5 respectively to achieve FLR equivalent to BER 1E-15. SNR is 27.9dB and 18.8dB (9.1dB higher for PAM8). Note this still assumes a FEC scheme for PAM8 has more latency and complexity.
- PAM8 results in higher DFE complexity.



PAM4 EYE



PAM8 EYE

# PAM4 and PAM8 Performance Comparison

- Maximum SNR at decision point can be computed by Salz SNR, which is:

$$\begin{aligned} SNR_{salz} &= 10 \cdot \log_{10} \left\{ \exp \left[ \frac{1}{FN} \int_0^{FN} \ln \left( 1 + \frac{S(f)}{N(f)} \right) \cdot df \right] \right\} \\ &\approx \frac{1}{FN} \int_0^{FN} 10 \cdot \log_{10} \left( \frac{S(f)}{N(f)} \right) \cdot df \\ &= AVG_{0 \leq f \leq FN} [SNR_{dB}(f)] \\ &\approx TX\ SNR - IL_{NY} / 2 \\ &= PT / (2N_0) - IL_{NY} / 2 \end{aligned}$$

where  $FN$  is Nyquist Frequency,  $P$  is TX signal power,  $IL_{NY}$  is insertion loss at  $FN$ . For simplicity, system noise is assumed to be AWGN, and channel is assumed to be dielectric loss (linear phase) dominant. For PAM8,  $T$  and  $IL_{NY}$  are both 2/3 of PAM4.

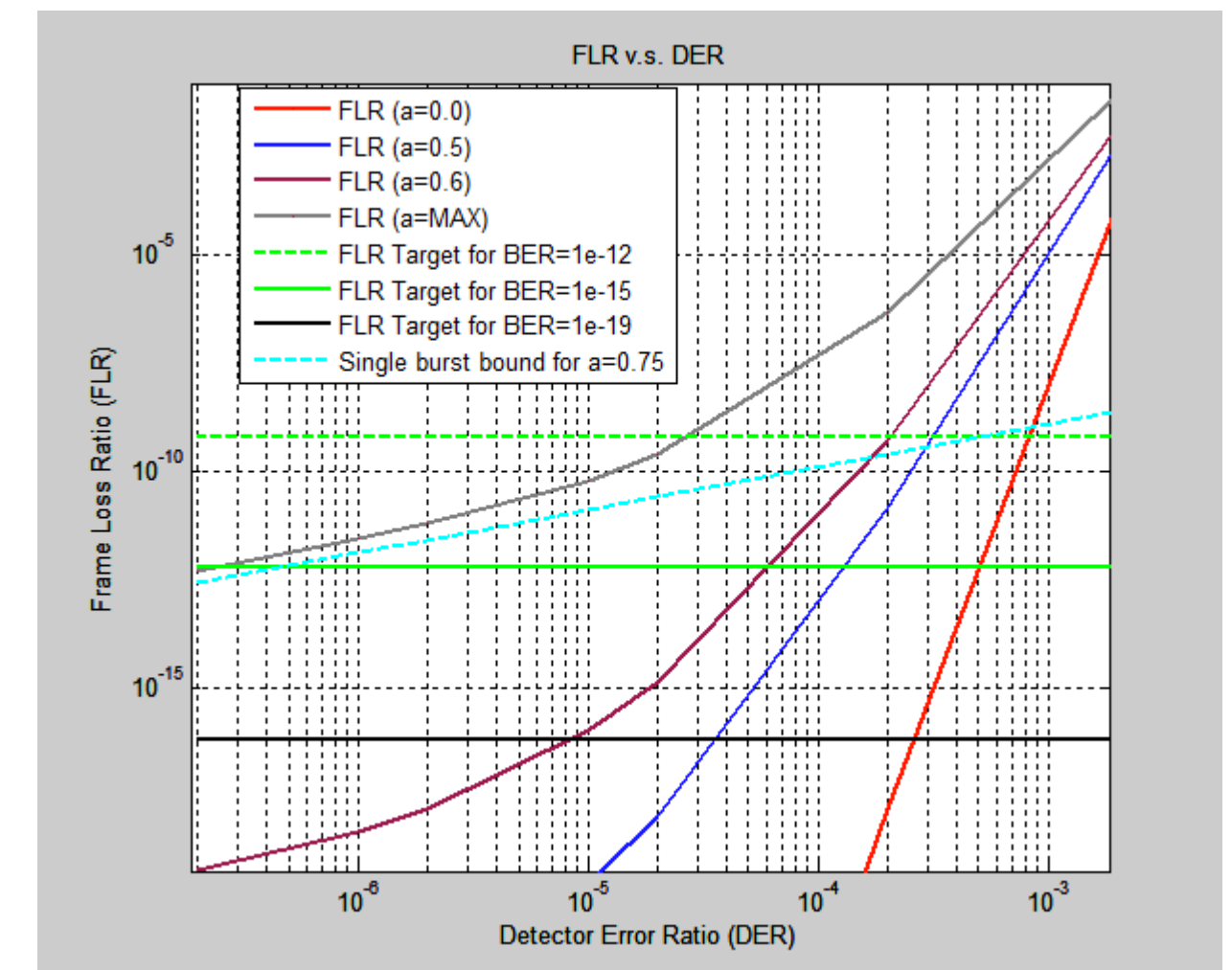
- PAM4 performs better for channels with IL less than 50.8dB at PAM4 Nyquist frequency. Considering skin loss, PAM4 performs better on even higher loss channel.

# DER Requirement and FEC

- For PAM4 modulation with KP4 FEC, worst DFE error propagation rate ('a') is 0.75. In this case, DER needs to be  $2.9\text{E-}5$  to achieve frame loss ratio equivalent to BER  $1\text{E-}12$  (FLR= $6.2\text{E-}10$ ). Raw BER requirement is  $5.8\text{E-}5$ .
- If DFE error propagation rate can be limited to 0.6, DER and BER requirement can be relaxed to  $2.1\text{E-}4$  and  $3.2\text{E-}4$ .
- Raw BER requirement needs to be lower (shared) if there are multiple links.
- If  $1\text{E-}15$  post FEC BER is required for some applications, burst error penalty is very high and needs to be controlled.

BER Target	FLR	a=0.75	a=0.6	a=0
1E-12	6.2E-10	2.9E-5	2.1E-4	7.6E-4
1E-15	6.2E-13	2.5E-7	6.0E-5	5.0E-4

DER Requirement



KP4 FEC Performance for PAM4

# DER Requirement and Interleaved FEC

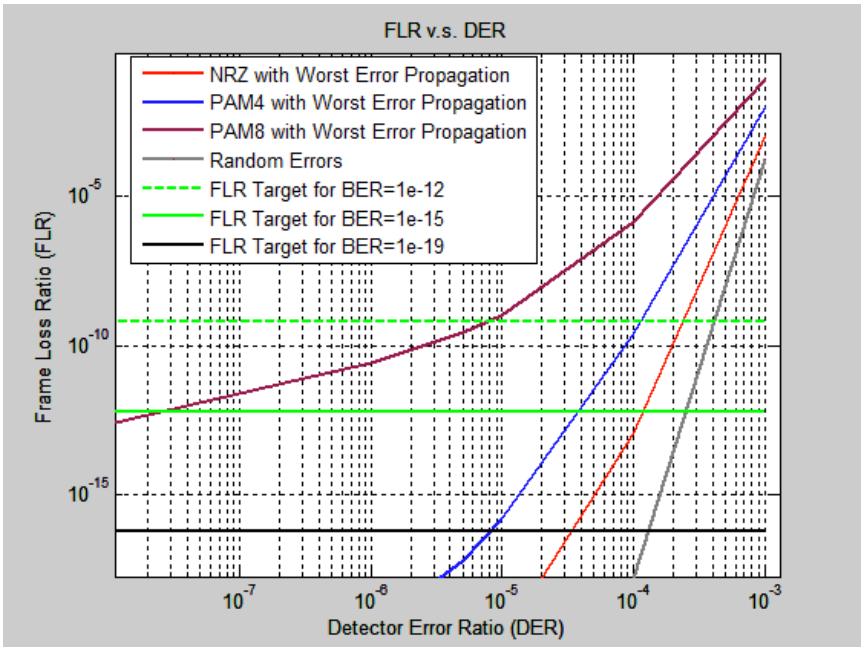
- Considering 1+D precoder is only effective on certain burst patterns, symbol interleaving is more reliable to treat burst errors.
- Assuming no interleaving for NRZ, 2-way interleaving for PAM4, 3-way interleaving for PAM8, KP4 FEC net coding gain is much less for PAM8 than NRZ and PAM4. 3-way interleaving also results in longer latency and higher complexity.
- Lane multiplexing schemes are not decided and may further degrade FEC coding gain to some extent.
- Preliminary simulation results in the following slides indicate PAM4 DER requirement is reasonable. PAM8 needs a stronger FEC and/or THP if 1E-15 is required for some applications.

BER Target	FLR	NRZ	PAM4	PAM8
1E-12	6.2E-10	2.3E-4	1.1E-4	7.8E-6
1E-15	6.2E-13	1.2E-4	3.8E-5	2.6E-8

## DER Requirement for Interleaved FEC

Interleave	no	2-way	3-way
100GE FEC Latency	110ns	160ns	210ns

FEC Latency

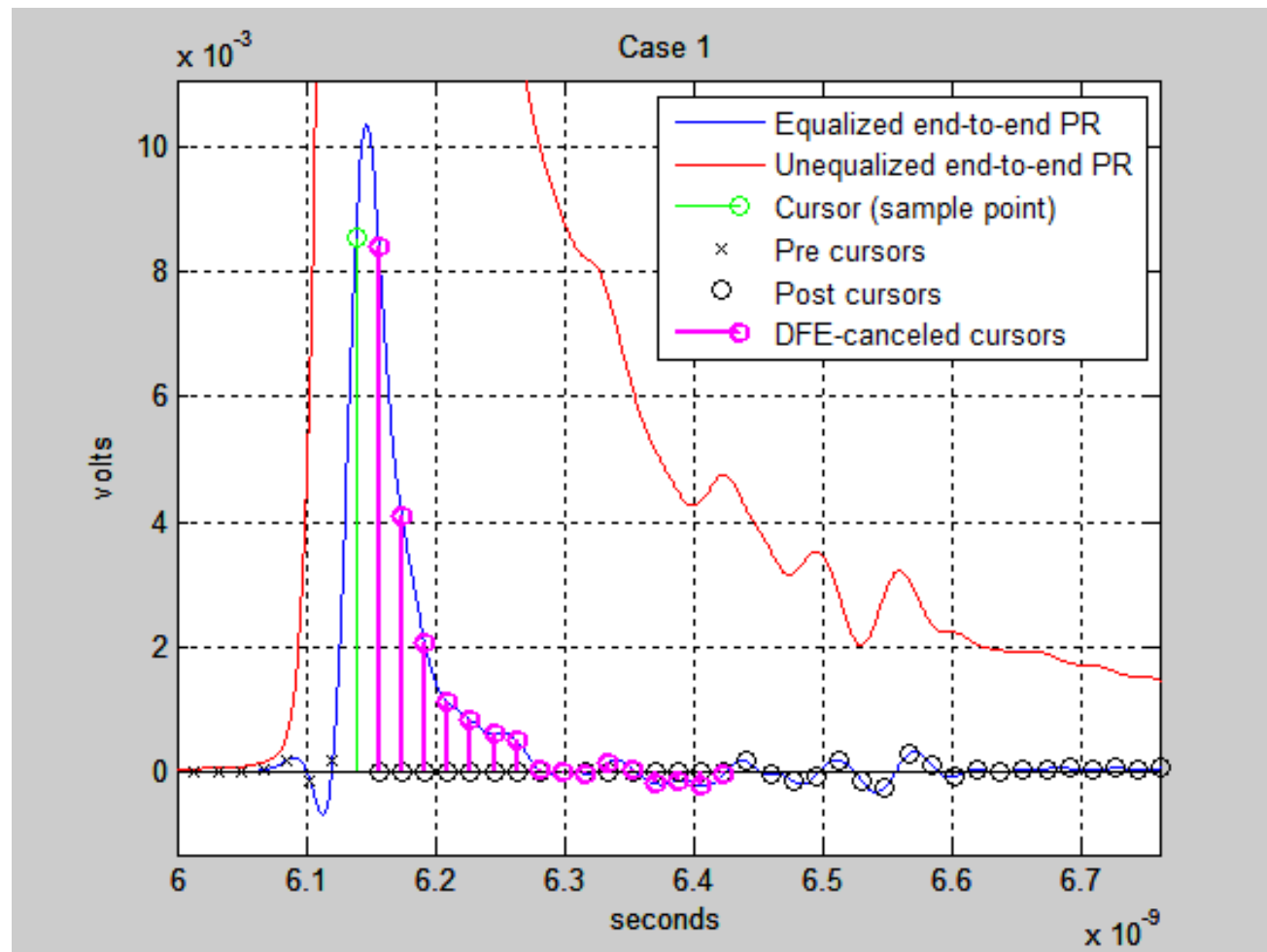


## Interleaved KP4 FEC Performance

No interleaving for NRZ,  
2-way interleaving for PAM4,  
3-way for PAM8



# Power Challenge of 100Gb/s LR SERDES



Single Bit Response of a channel intended for 100Gb/s

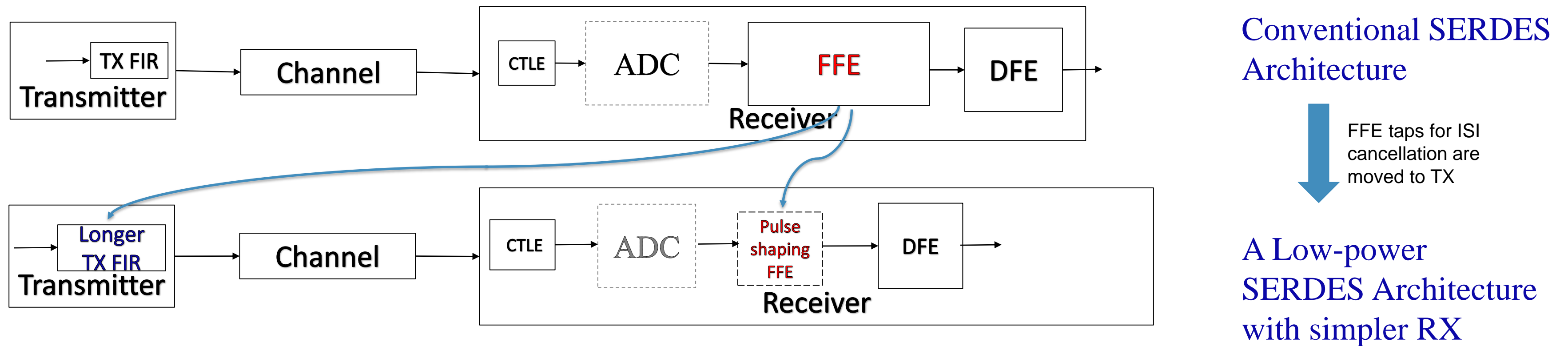
- Given the same channel, reflections may appear on double number of UI's because of double Baud Rate.
- FFE or DFE is commonly used for equalization and consumes a big portion of SERDES power (usually 25% to 50% depending on architecture). FFE or DFE needs double number of taps and double throughput compared to 50Gb/s. Power of RX FFE or DFE theoretically will be up to 4x on the same process node!
- Because throughput or bandwidth doubles, power of other major components (ADC, TX, CTLE) theoretically double as well.

**For a switch ASIC with 128 or 256 ports, this power increase is significant!**

**Solutions need to be found! Lower power SERDES Architecture or better channels?**

**Maybe both.**

# Lower-Power 100Gb/s Architecture Opportunity

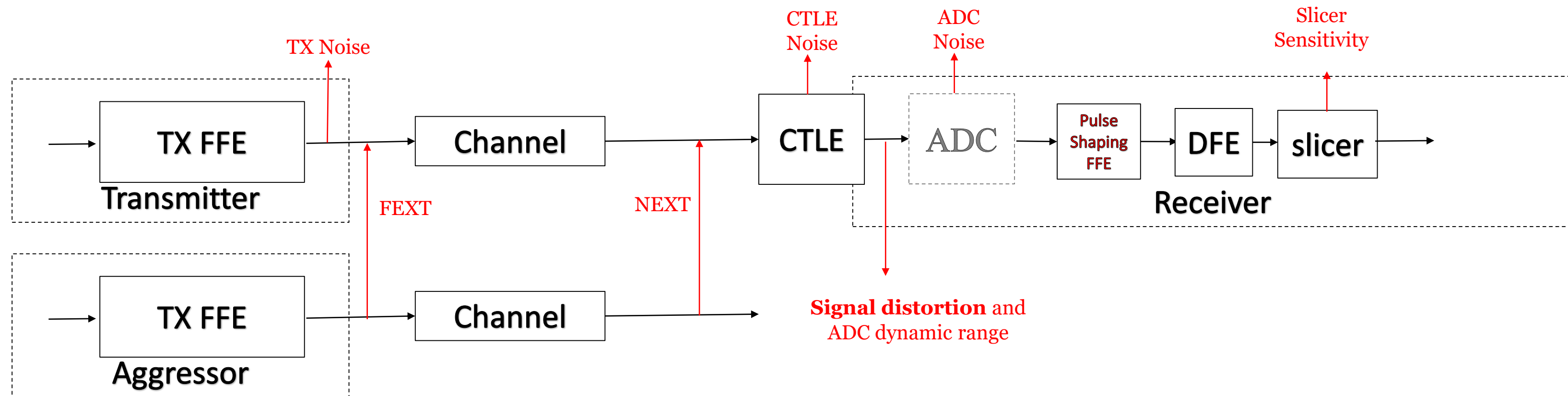


- Proposed SERDES moves long FFE for ISI cancellation to TX. Therefore, receiver can be much simpler and easier. For example, CTLE, 1-tap DFE, and possibly a few taps FFE (depending on channel profile and CTLE structure) to deal with material loss.
- TX FFE is much less expensive than RX FFE because input bit width is much less and multipliers can be avoided.
- ADC power can be reduced as well as dynamic range is reduced.
- TX-centric equalization is not new. It is commonly used to save receiver power, and manage interference. In SERDES case, TX FIR costs much less compared to RX FFE/DFE. Interoperation and test experience can be borrowed from these projects.

**About 30% SERDES power reduction compared to conventional architecture!**

# Noise and Distortion Analysis

Distortion, noise, complexity, and power caused by complex circuits shall be considered. Not just theoretical performance comparison of mathematical equalization models.



Noise and distortion sources

# Noise and Distortion Analysis cont.

## ➤ Signal:

- Same at slicer input if system is linear.

## ➤ Distortion and noise:

- Reflection ISI can be better cancelled because more FFE taps can be implemented on TX side with lower power.
- Easier RX and better **linearity**. For example, CTLE output signal dynamic range is smaller and less distortion. Important for PAM4 signal.
- ADC needs less dynamic range, and no noise enhancement by RX FFE.
- XTALK: Aggressors have lower PSD. Same XTALK impact from aggressors using the same structure.
- NEXT and CTLE noise are relatively boosted higher. The difference can be controlled by letting RX take care of material loss.

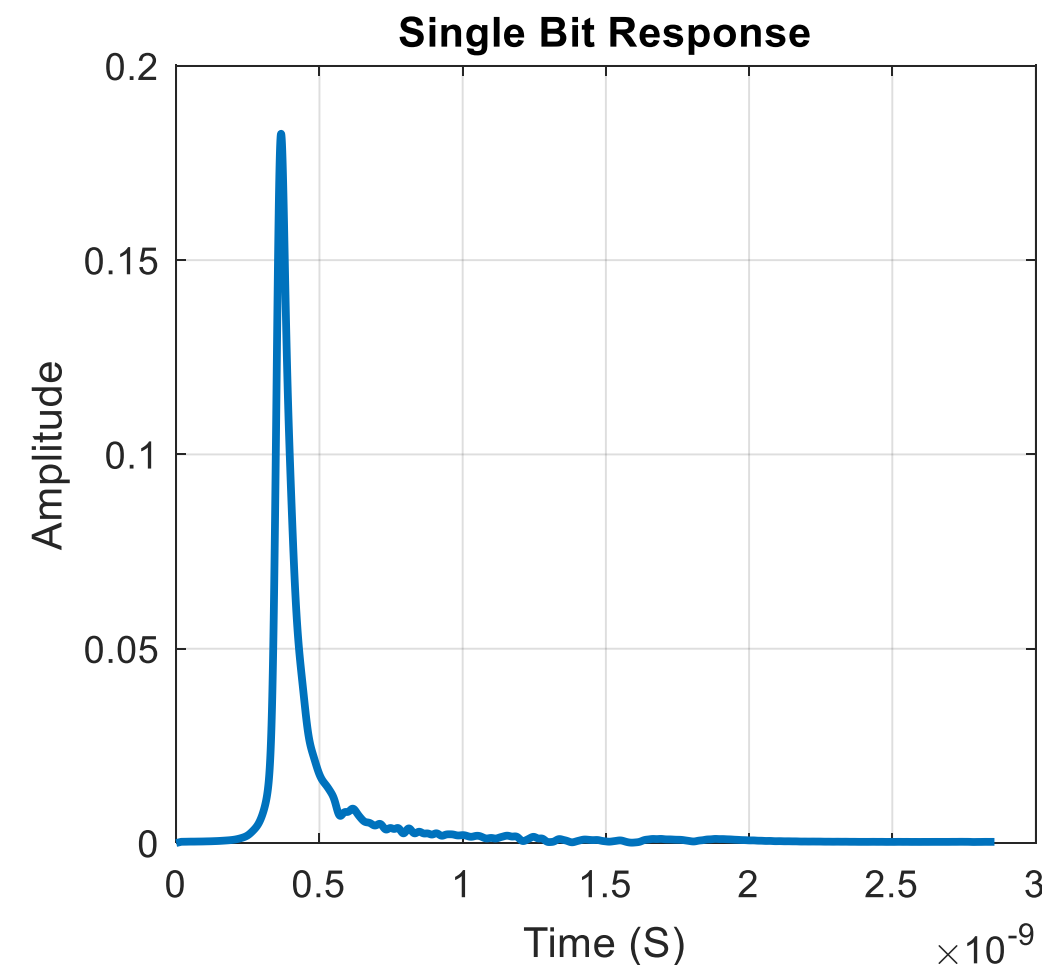
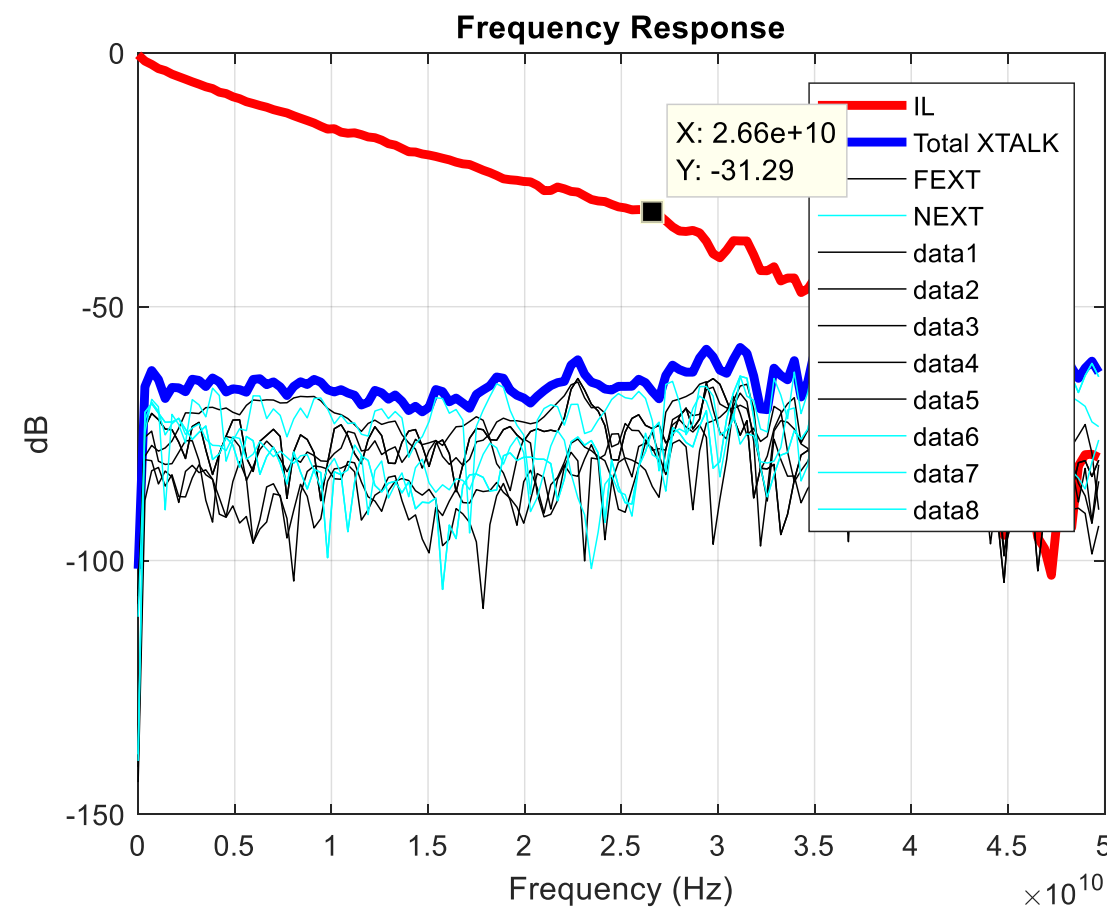
## ➤ Noise enhancement and distortion tradeoff is application dependent. The advantage of this “balanced” TX/RX equalization scheme is to cancel reflections and alleviate distortion with significantly less power.



# Simulation Setup

## ➤ Test Channel

- Latest orthogonal backplane channel model on IEEE 802.3 100GEL website (tracy\_100GEL\_03\_0118.pdf).
- 31dB IL at 26.5625GHz including packages and parasitic caps. Relatively “smooth”.
- 5 FEXT and 3 NEXT channels.



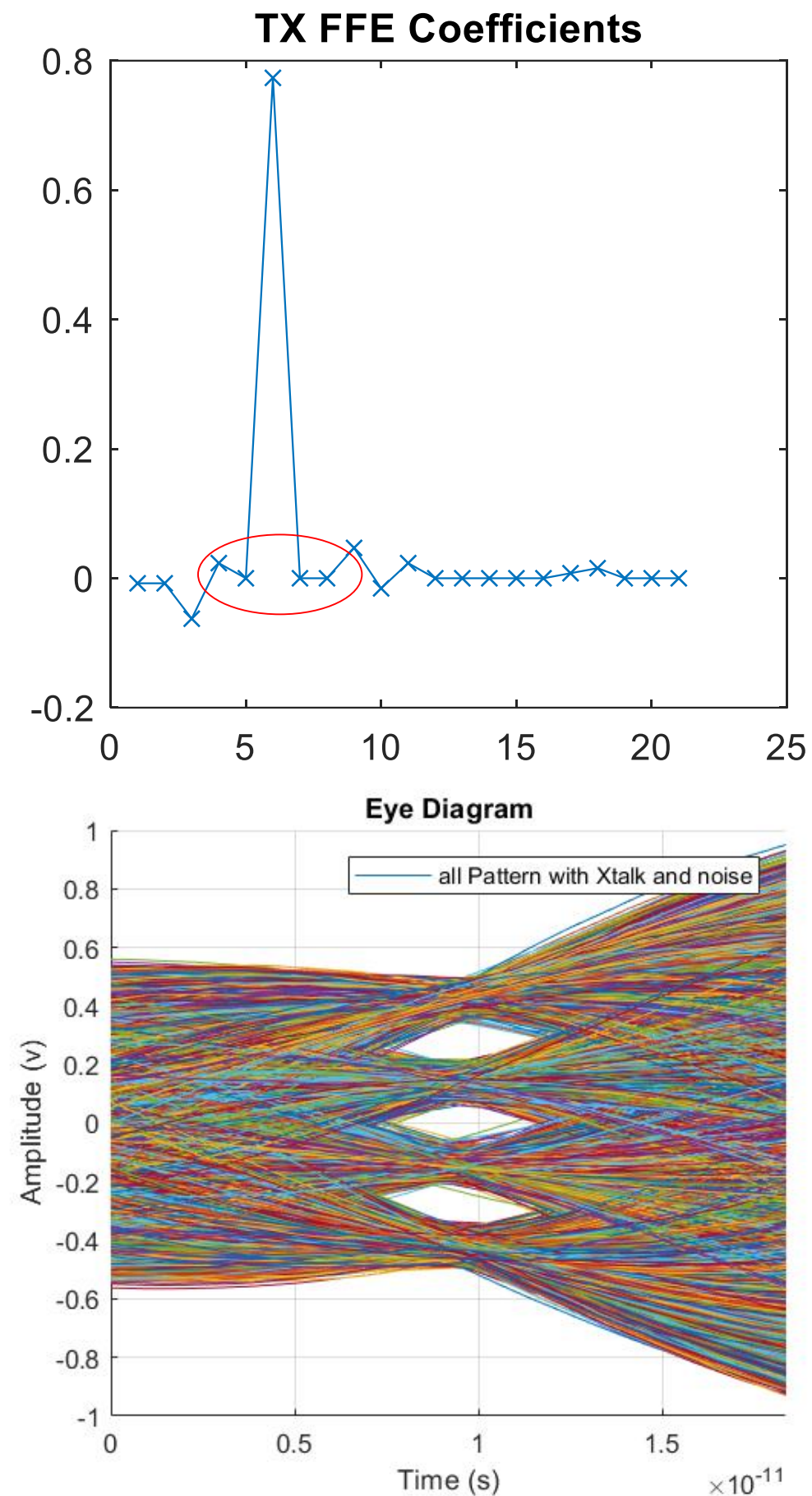
## ➤ TX SNDR 34dB

## ➤ Jitter: RJ 0.01 UI RMS, DJ: 0.04 UI p2p.

## ➤ Circuit noise and distortion are included.

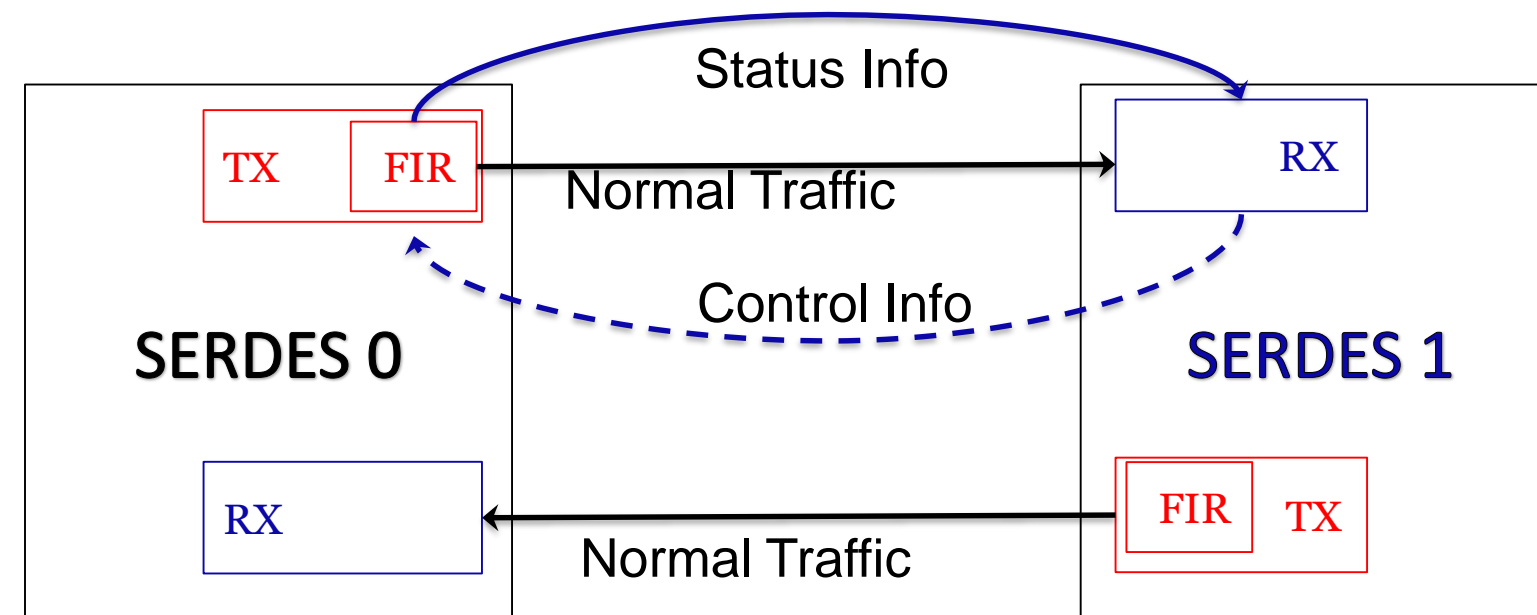
# Simulation Result

- TX/RX “balanced” Equalization:
  - TX: 22-tap FIR, 5 pre.
  - RX: CTLE, 5-tap FFE, 1-tap DFE
  - Main cursor weight is about 78%. Tuning algorithm can be further optimized.
  - BER is  $3E-7$ .
- For the simulated channel, power and BER of the proposed architecture are promising. For channels having longer reflections, TX FFE may need more taps.

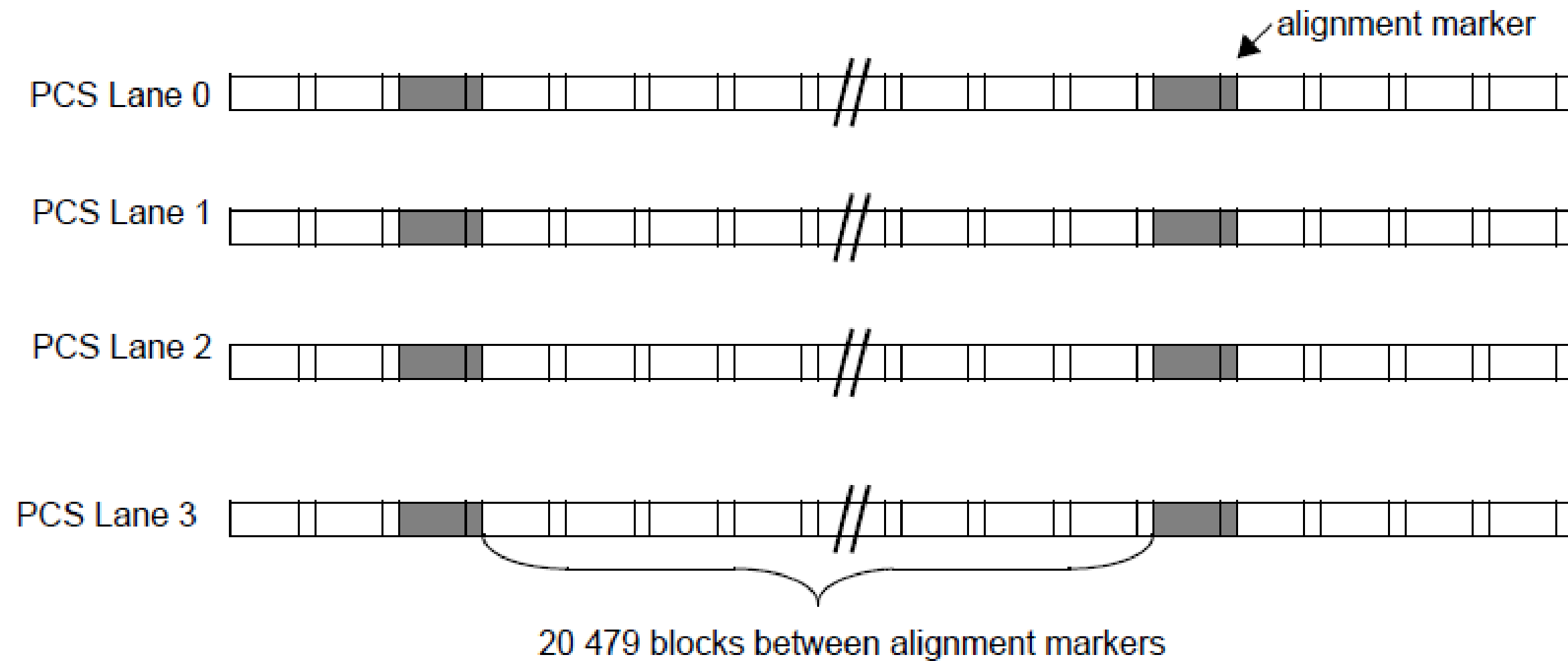


# TX FIR Real-Time Adaptation

- Startup training mechanism is defined in IEEE 802.3 Clause 94 and 136. The purpose is to adapt TX FIR.
  - Channels have big variation due to temperature or humidity.
  - More TX FIR taps are needed for 100G. More impact on performance.
  - TX FIR real-time adaptation is desired for optimal performance and simpler RX?
  - This type of adaptation rate can be low because channel variation is slow,
- **How to pass training information to remote TX during normal data traffic?**
- Training info include control and status and need to travel two directions.



# Finding Back Channel



- Alignment Marker is inserted for lane alignment and FEC boundary (e.g. IEEE802.3 clause 133), and is mandatory for links with PAM4 signaling.
- **Back channel mechanism: TX add status and control field (from local RX) into alignment marker. RX has a detection logic to lock to the alignment marker, and fetch status and control commands (for local TX).**



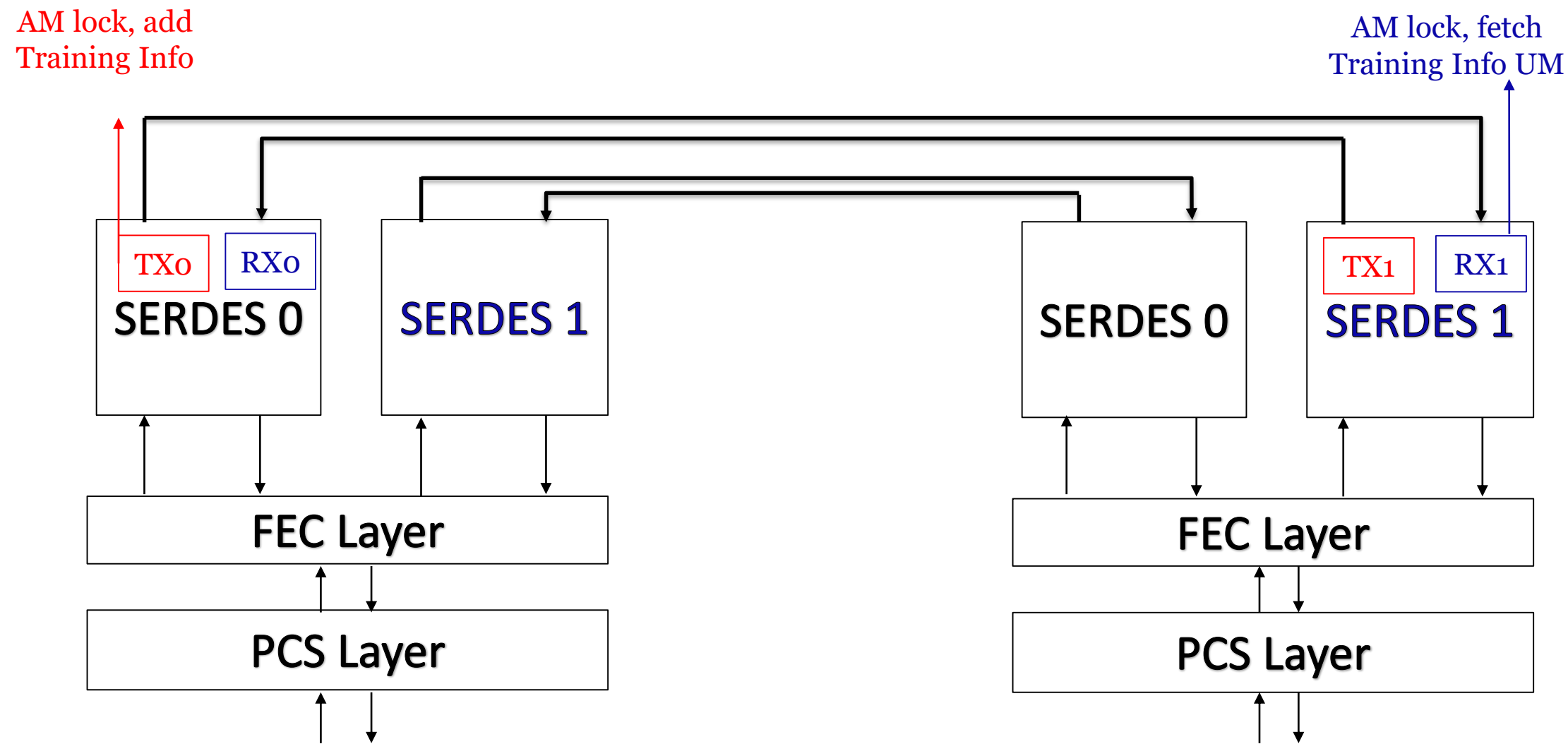
# TX FIR Update Rate

- AM spacing.

Speed	AM spacing (66b blocks)	Time interval between AM (us)
50GE	20480x4	104.86
100GE	16384x20	209.72
200GE	81920x4	104.86
400GE	163840x4	104.86

- Update rate is about 5000-10000 times per second, enough to track temperature/humidity variations.

# Back Channel Mechanism



Back Channel Diagram

- This AM lock is done in SERDES for repeater applications. This logic only locks to AM without doing alignment. Hardware cost is trivial.
- For implementations with FEC layer, logic could be shared.

# Training Info Field

- To get back channel, some bits in AM can be reserved or reused for each FEC lane.
- For example, reserve some bits in 2 FEC lanes as shown in the following figure.
- Reliability can be guaranteed by error detection protocols.
- RX knows whether training info should be expected in AM during frame training or through MDIO.

FEC Lane	Reed Solomon Symbols					
Lane 0	amp_tx_0(0:63)	amp_tx_4(0:63)	amp_tx_8(0:63)	amp_tx_12(0:63)	command field	
Lane 1	amp_tx_1(0:63)	amp_tx_5(0:63)	amp_tx_9(0:63)	amp_tx_13(0:63)	status field	
Lane 2	amp_tx_2(0:63)	amp_tx_6(0:63)	amp_tx_10(0:63)	amp_tx_14(0:63)		
Lane 3	amp_tx_3(0:63)	amp_tx_7(0:63)	amp_tx_11(0:63)	amp_tx_15(0:63)		

A possible approach of back channel bits allocation

# TX Training Time

- Current TX FIR training updates only one coefficient per training frame.
- More TX FIR taps are needed for 100G and results in more TX training work.
- Can longer training time be tolerated by upper layer?
- Can we update multiple coefficients simultaneously to speedup?
  - Need to extend control/status field structure to have dedicated bits for each coefficient.
- It may be useful to add status information, such as the number of unused drivers, step size, and coefficient weight.

Table 136-9—Control field structure

Bit(s)	Name	Description
15:14	Reserved	Transmit as 0, ignore on receipt
13:12	Initial condition request	13 12 1 1 = Preset 3 1 0 = Preset 2 0 1 = Preset 1 0 0 = Individual coefficient control
11:10	Reserved	Transmit as 0, ignore on receipt
9:8	Modulation and precoding request	9 8 1 1 = PAM4 with precoding 1 0 = PAM4 0 1 = Reserved 0 0 = PAM2
7:5	Reserved	Transmit as 0, ignore on receipt
4:2	Coefficient select	4 3 2 1 1 0 = $c(-2)$ 1 1 1 = $c(-1)$ 0 0 0 = $c(0)$ 0 0 1 = $c(1)$
1:0	Coefficient request	1 0 1 1 = No equalization 1 0 = Decrement 0 1 = Increment 0 0 = Hold



# Summary

- 100Gb/s PAM4 SERDES is desired for higher speed interconnect and being shown on silicon.
- Two modulation schemes are compared. PAM4 is preferable than PAM8 considering joint performance of FEC and SERDES.
- DER requirement for interleaved KP4 FEC is studied. With the development of channels and SERDES, there will be more information whether stronger FEC is needed.
- SERDES power may dramatically increase due to equalization challenge and throughput of 100Gb/s electrical link, and result in significant ASIC power increase.
- A low-power architecture opportunity for 100Gb/s LR SERDES :
  - **A standard supporting TX/RX “balanced” equalization will enable remarkable SERDES power reduction!**
- Real-time TX training and faster adaptation mechanism are introduced for robust SERDES performance.

*Thanks!*