agere systems

Exceeding the Speed Limit

Challenges and solutions to the support of serial 10 Gigabit Ethernet over backplane interconnects

Adam Healey

Chair, IEEE P802.3ap Task Force

Consulting Member of Technical Staff, Agere Systems

March 10, 2006

Before we begin…

- Per IEEE-SA Standards Board Operations Manual, January 2005:
  - At lectures, symposia, seminars, or educational courses, an individual presenting information on IEEE standards shall make it clear that his or her views should be considered the personal views of that individual rather than the formal position, explanation, or interpretation of the IEEE.

- Note:
  - Currently, IEEE P802.3ap is under review per the IEEE 802.3 Working Group ballot process and is subject to change.
  - The work in this presentation is per IEEE P802.3ap Draft 2.1.

Agenda

- Introduction to Backplane Ethernet
- Backplane Architecture and Reference Model
- 10GBASE-KR
- Forward Error Correction for 10GBASE-KR
- Closing Remarks

Backplane Ethernet Profile

- "Ethernet in a box"
- Target Applications
  - Wireline and wireless access equipment
  - Blade servers
  - Enterprise switching
- Leverages field proven Ethernet controller and switching IP
- Reduces solution cost and complexity
  - Leverages Ethernet economies of scale
  - Flattens the backplane fabric eliminating encapsulation protocols and associated processing
  - Standard enables multi-vendor interoperability and facilitates use of COTS devices

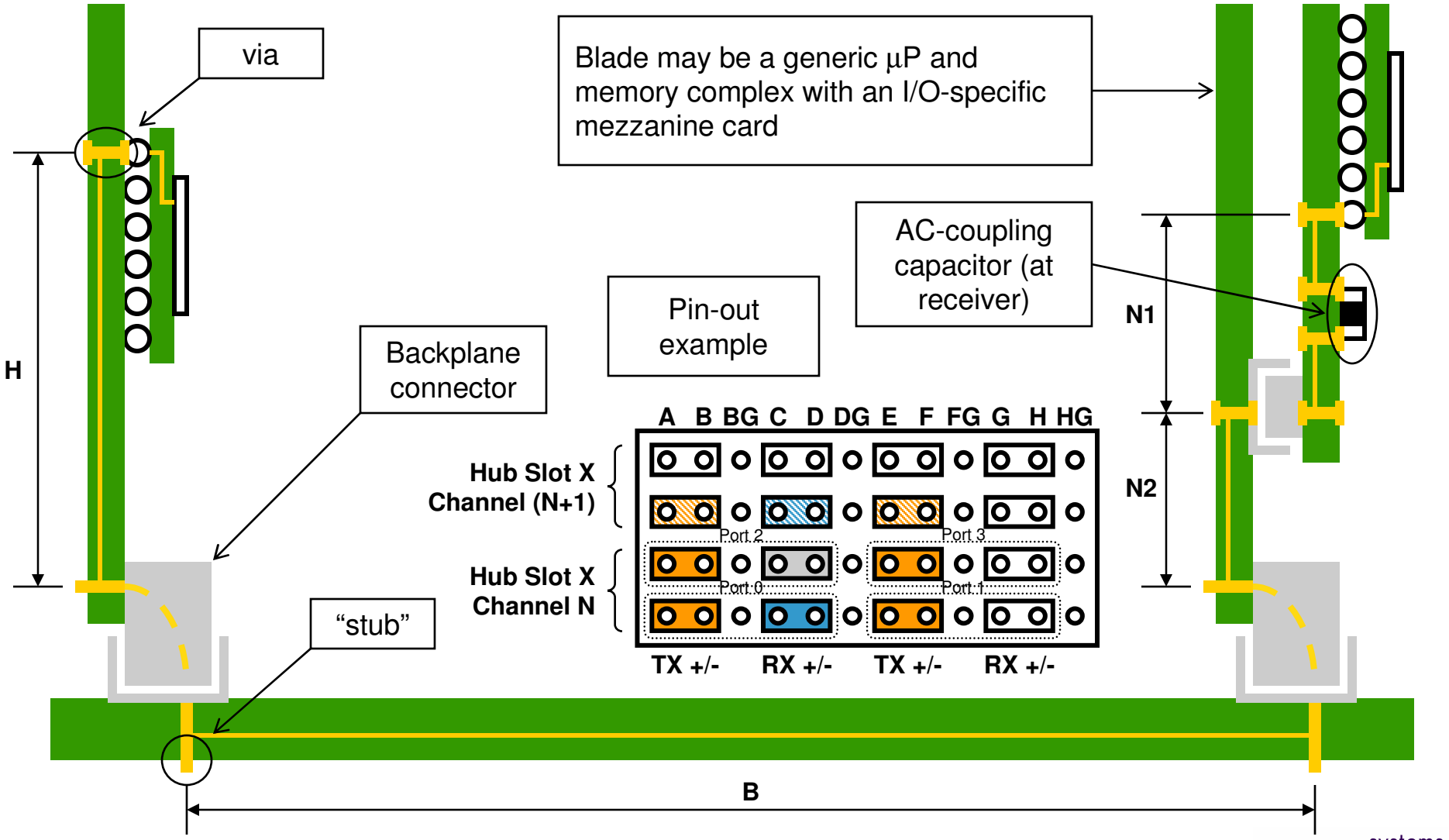
Specification Methodology

- Supply a device specification and not a backplane specification
- Be application agnostic (do no specify the backplane connector)
- Supply informative recommendations to assist backplane designers in identifying backplane channels that are interoperable with “Backplane Ethernet compliant” devices

Agenda

- Introduction to Backplane Ethernet
- Backplane Architecture and Reference Model
- 10GBASE-KR
- Forward Error Correction for 10GBASE-KR
- Closing Remarks

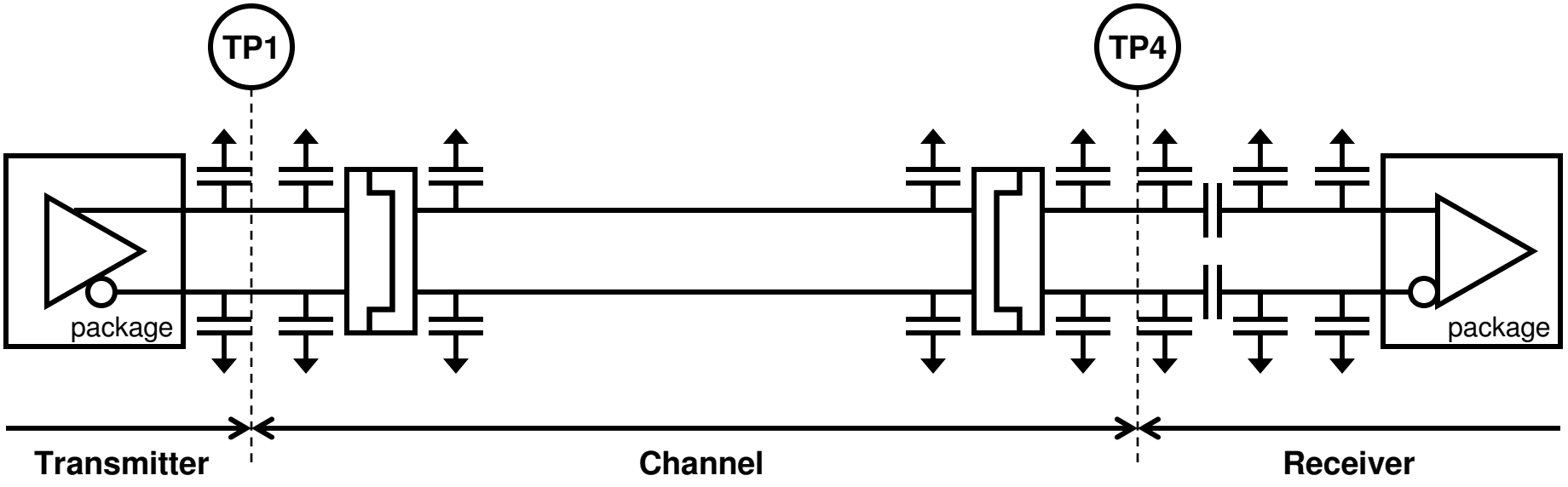
Backplane Architecture and Topology



agere systems

Reference Model

- The transmitter and receiver blocks include all off-chip components associated with the respective block
  - Example, external AC-coupling capacitors, if required, are included in the receiver block
- Channel consists of line card and backplane traces plus associated backplane and mezzanine connectors



agere systems

Backplane Ethernet Objectives

- Support up to 1 m differential traces, including two connectors, on *improved* FR-4 printed circuit boards
- Support a BER of 10^{-12} or better

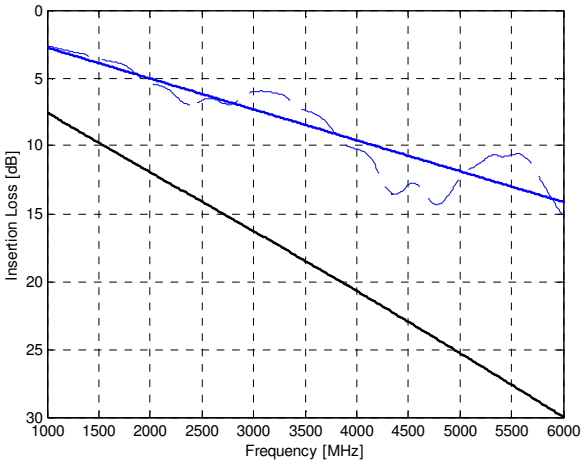
Description	N1 (mm)	N2 (mm)	B (mm)	H (mm)	Total (mm)	No. Connectors	AC / DC Coupling
AdvancedTCA							
Example (dual-star)[1]	0	102	244	102	447	2	AC
Full mesh[2]	0	127	533	127	787	2	AC
Blade Server							
Proposed worst-case[3]	76	102	533	127	838	3	AC
Switch / Router							
2 to 3 chassis/rack[4]	0	152	559	305	1016	2	AC or DC
5 to 8 chassis/rack[4]	0	127	432	229	787	2	AC or DC

[1] kundu_01_0504, [2] PICMG 3.0 R2.0, March 18, 2005, [3] koenen_01_0504, [4] goergen_01_0304

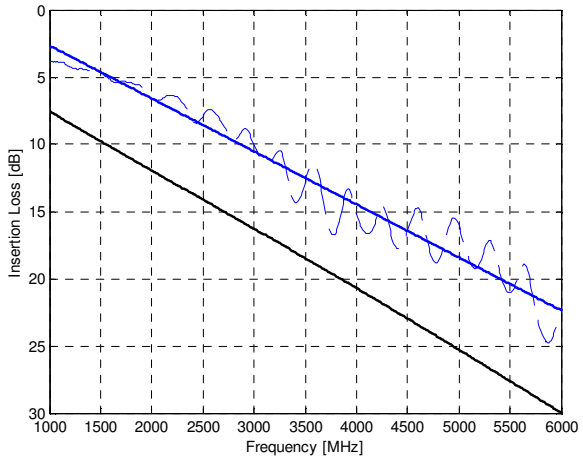
agere systems

Fitted Attenuation

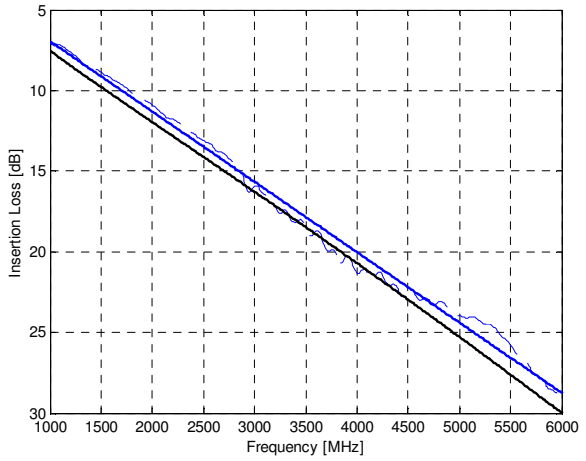
- Defined to be the least-mean squares line fit to the insertion loss data within the frequency range f_1 and f_2
  - The values of f_1 and f_2 are dependent on the PHY type of interest
- Fitted attenuation is compared to a limit based on a 1 m differential trace (W = 6 mil) on an *improved* FR-4 printed circuit board
- In effect, constrains the dielectric loss-length product of the channel



Example #1



Example #2

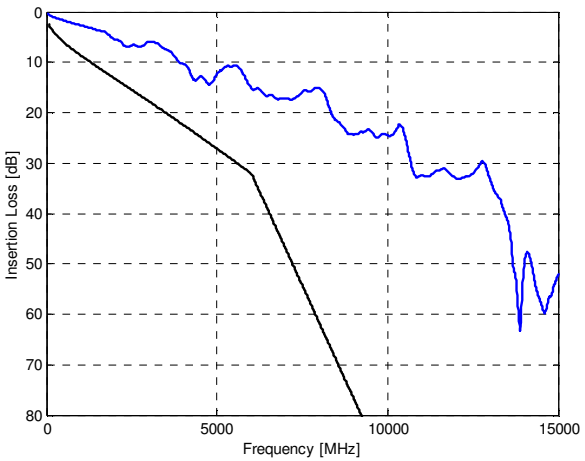


Example #3

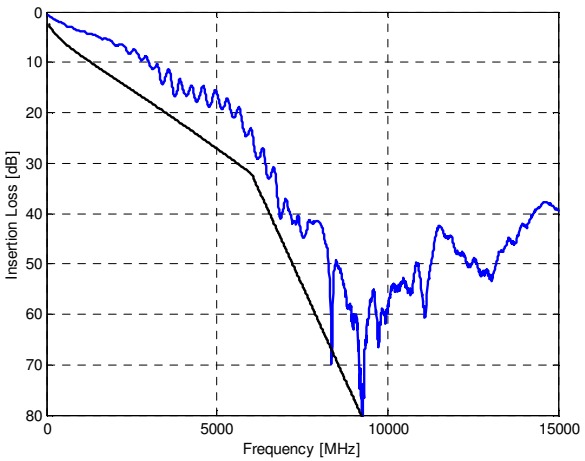
agere systems

Insertion Loss Limits

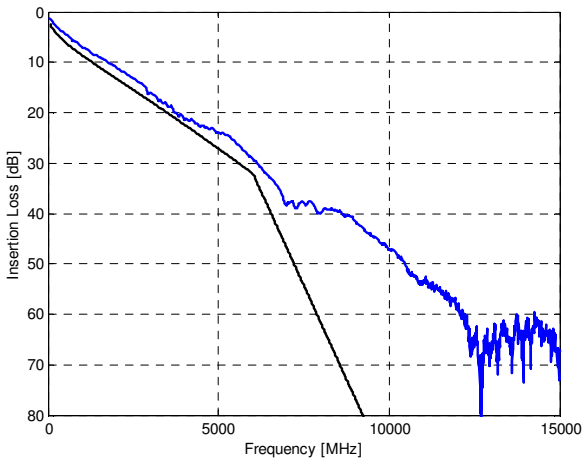
- Discourages encroachment of stub-related resonances into critical frequencies for the PHY type of interest
- Based on the fitted attenuation limit with allowances for *passband ripple* and stub-related resonances



Example #1



Example #2

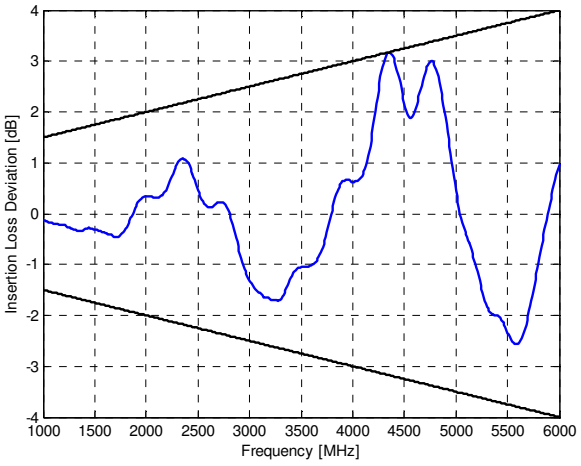


Example #3

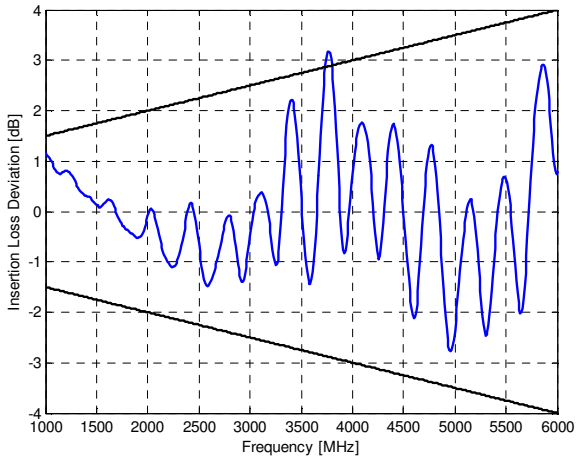
agere systems

Insertion Loss Deviation

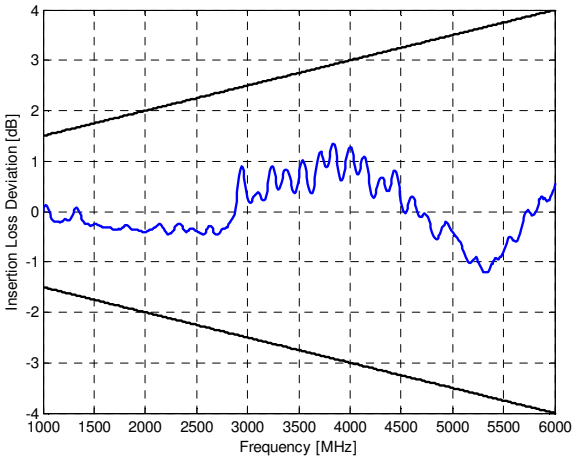
- Defined to the be the difference between the channel insertion loss and the fitted attenuation in the frequency range f_1 to f_2
- In effect, constrains *passband ripple* due to impedance mismatches in the transmission path



Example #1



Example #2

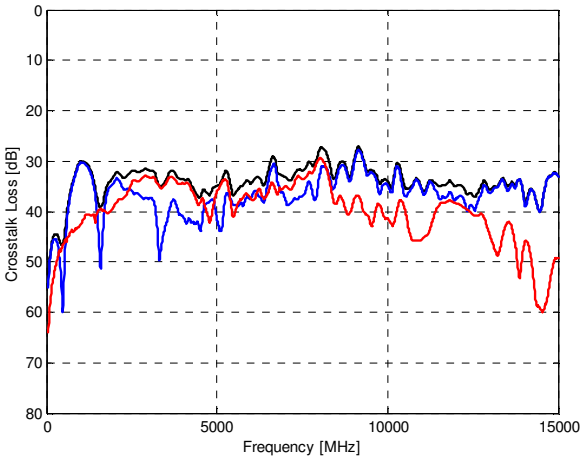


Example #3

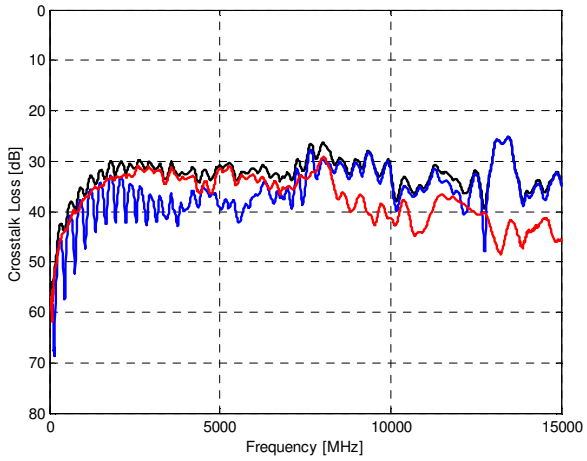
agere systems

Crosstalk

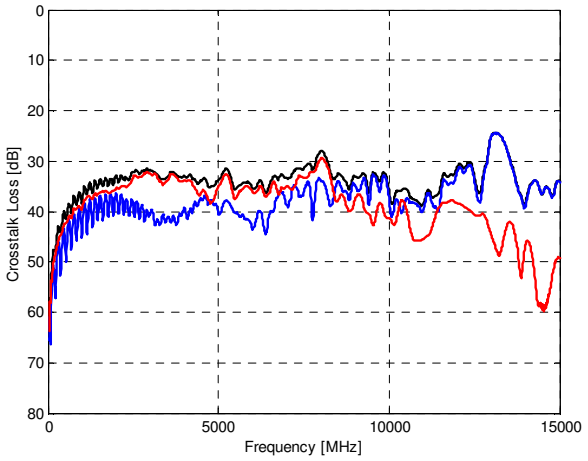
- Total crosstalk is defined to the be the power sum of the individual aggressors as measured at TP4
- Includes near-end crosstalk (NEXT) and far-end crosstalk (FEXT)
- Assumes the aggressors and victim are asynchronous and (or) uncorrelated
- Assumes the aggressors and victim are driven by identical sources



Example #1



Example #2

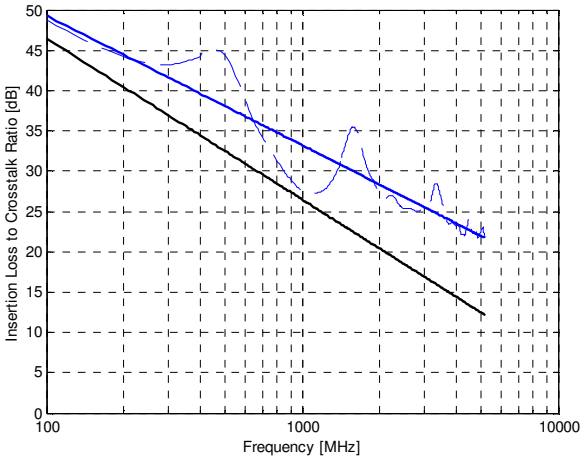


Example #3

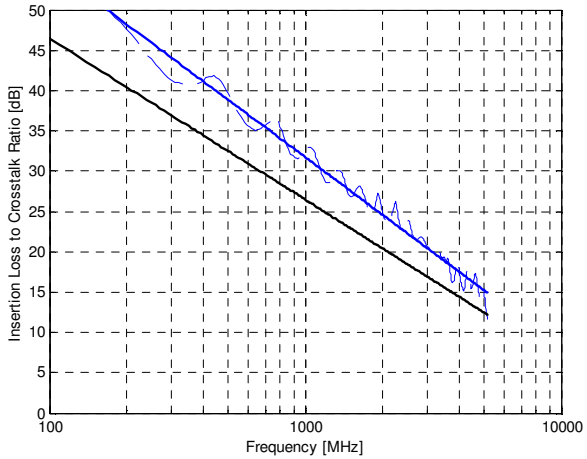
agere systems

Insertion Loss to Crosstalk Ratio (ICR)

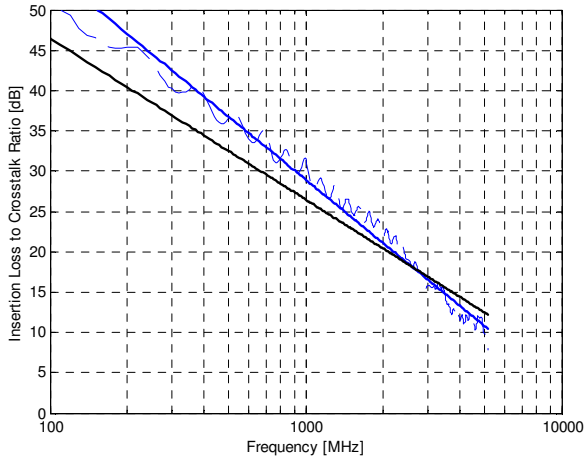
- Defined to be the least-mean squares line fit to the difference of the total crosstalk loss and the insertion loss (both in dB) in the frequency range f_a to f_b
- Limits the total crosstalk seen at TP4
- Lower loss links are allowed higher crosstalk and vice versa



Example #1



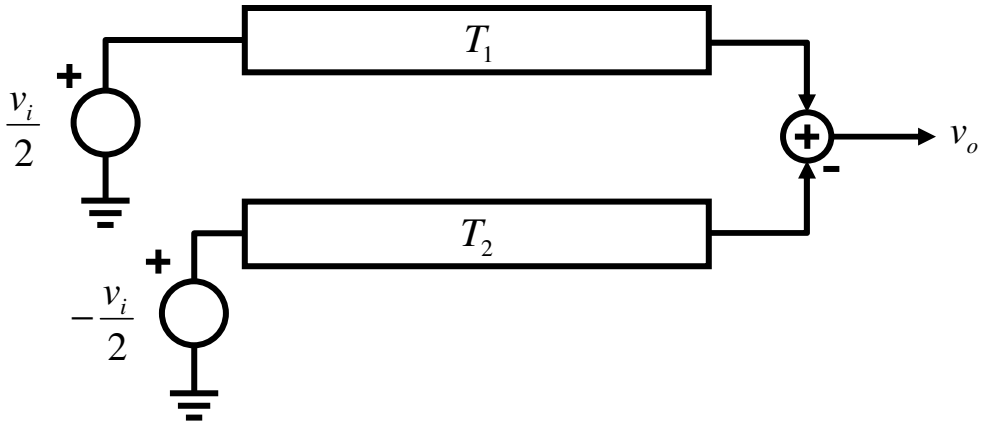
Example #2

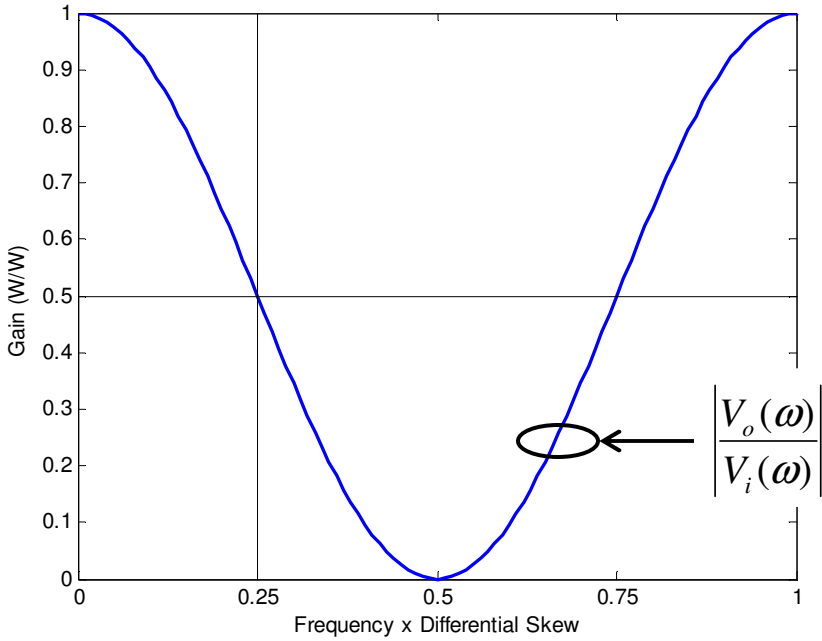


Example #3

agere systems

Differential Skew





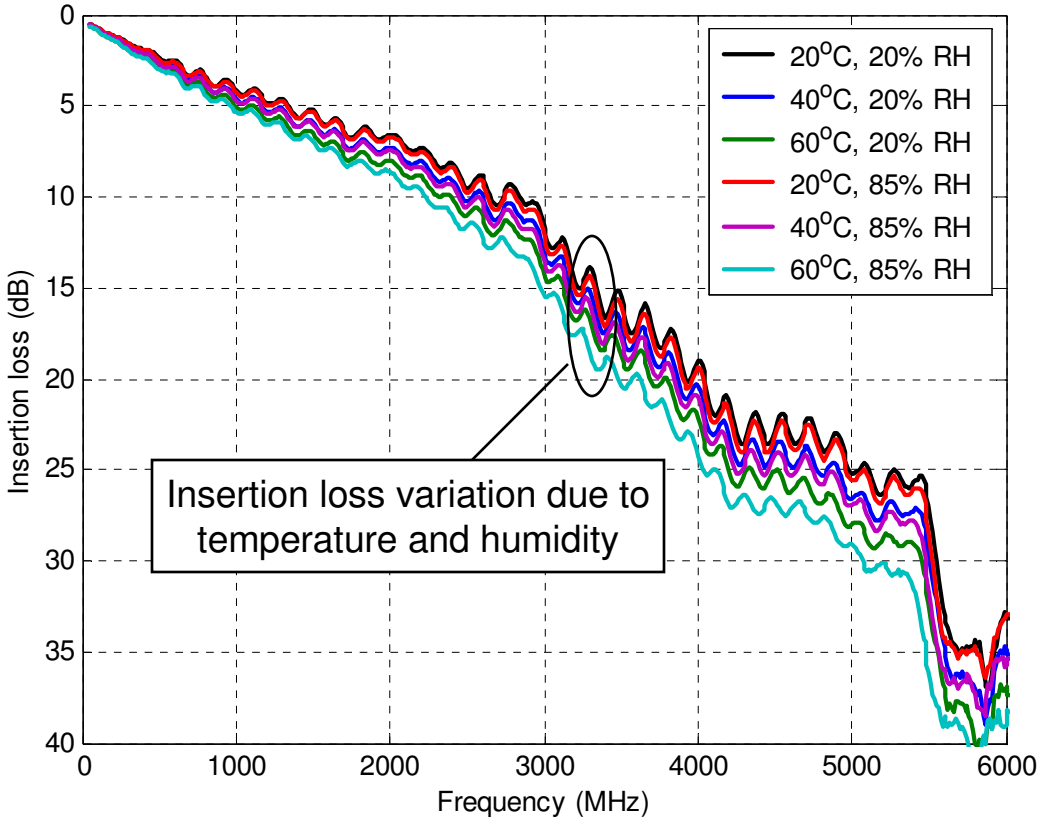
$$\left|\frac{V_o(\omega)}{V_i(\omega)}\right|^2 = \frac{1}{2}\left(1+\cos\left(\omega\left(T_2-T_1\right)\right)\right)$$

- To isolate the impact of skew, render the <p> and <n> sides of the differential pair as delay T_1 and T_2 respectively
- The skew is the magnitude of the difference between the two delays, $|T_1 - T_2|$
- The differential skew from TP1 to TP4 is recommended to be less than 24 ps (10GBASE-KR)

agere systems

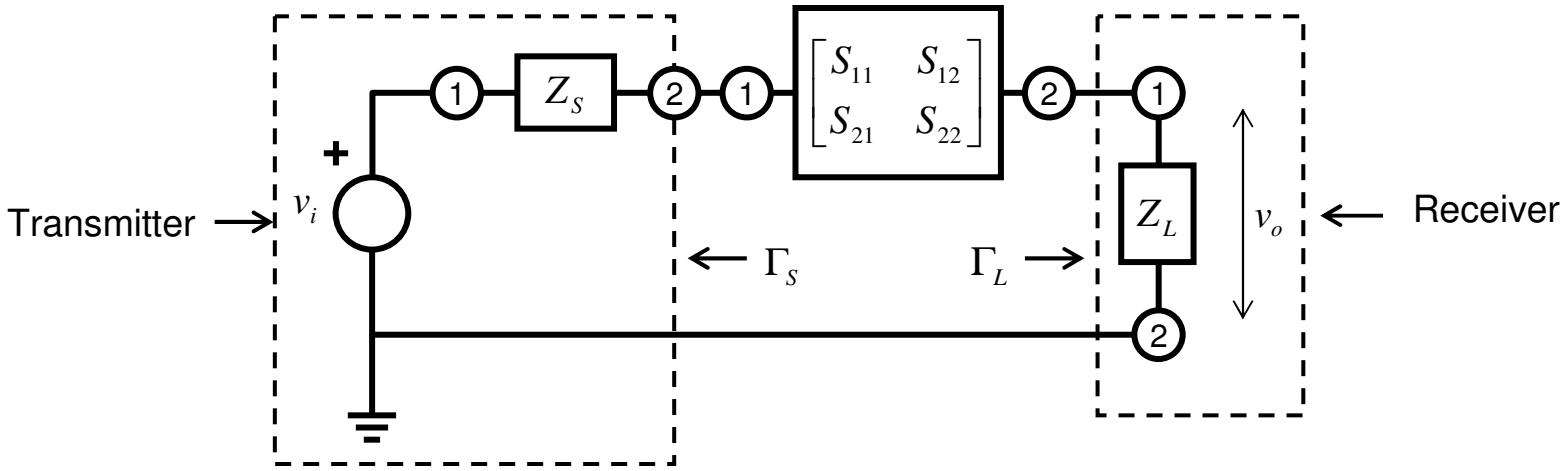
Environmental Variation

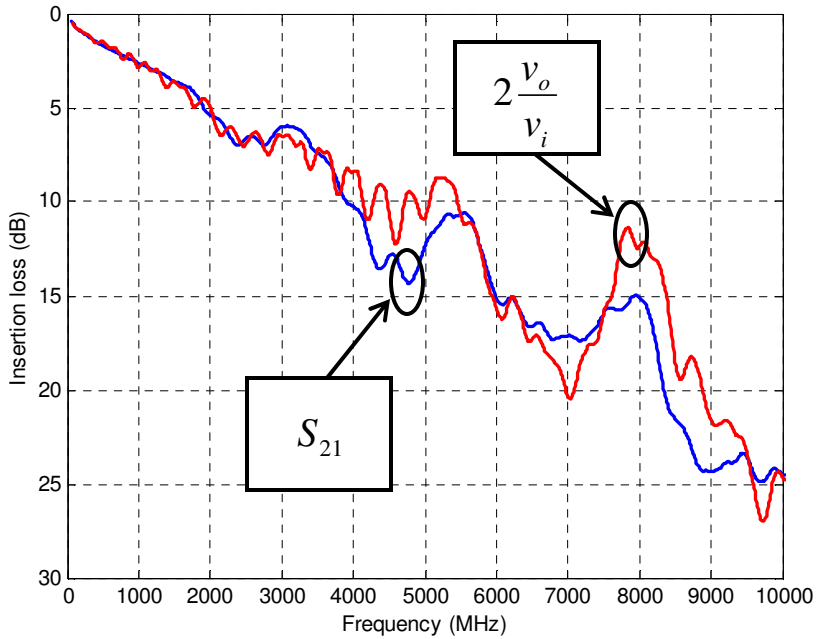
- Channel recommendations are to be satisfied over the full range of system operating conditions



agere systems

Impact of Device Terminations





$$\frac{v_o}{v_i} = \frac{S_{21}}{2}\left(\frac{(1+\Gamma_L)(1-\Gamma_S)}{1-\Gamma_S S_{11}-\Gamma_L S_{22}-\Gamma_S\Gamma_L S_{12}S_{21}+\Gamma_S\Gamma_L S_{11}S_{22}}\right)$$

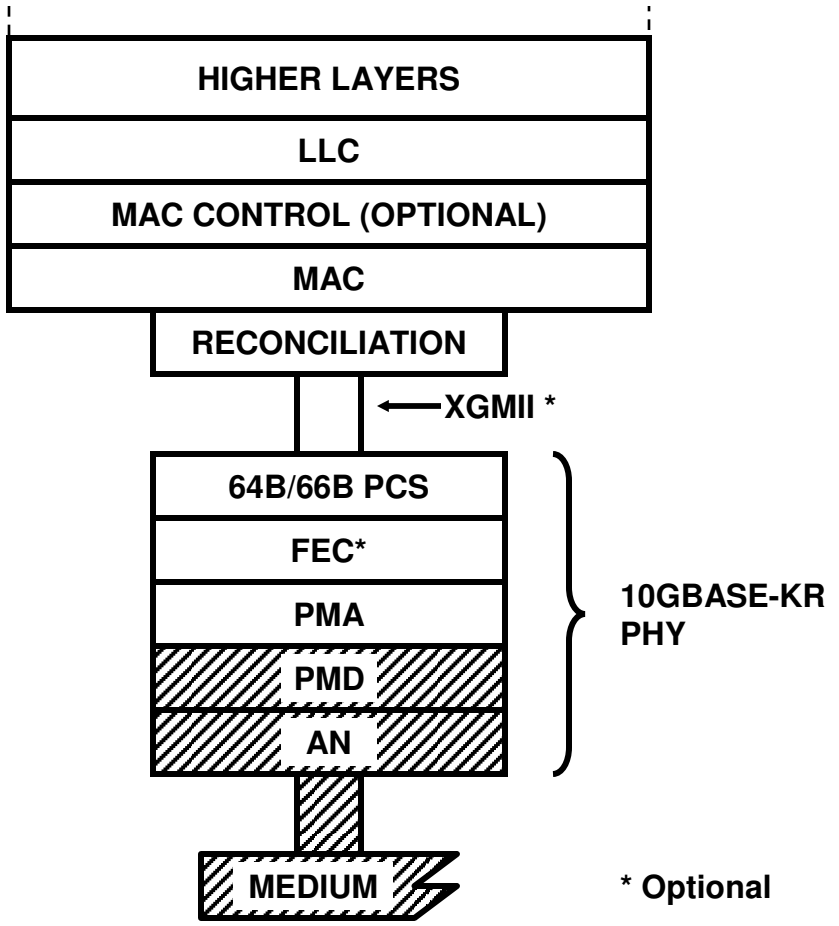
agere systems

Agenda

- Introduction to Backplane Ethernet
- Backplane Architecture and Reference Model
- 10GBASE-KR
- Forward Error Correction for 10GBASE-KR
- Closing Remarks

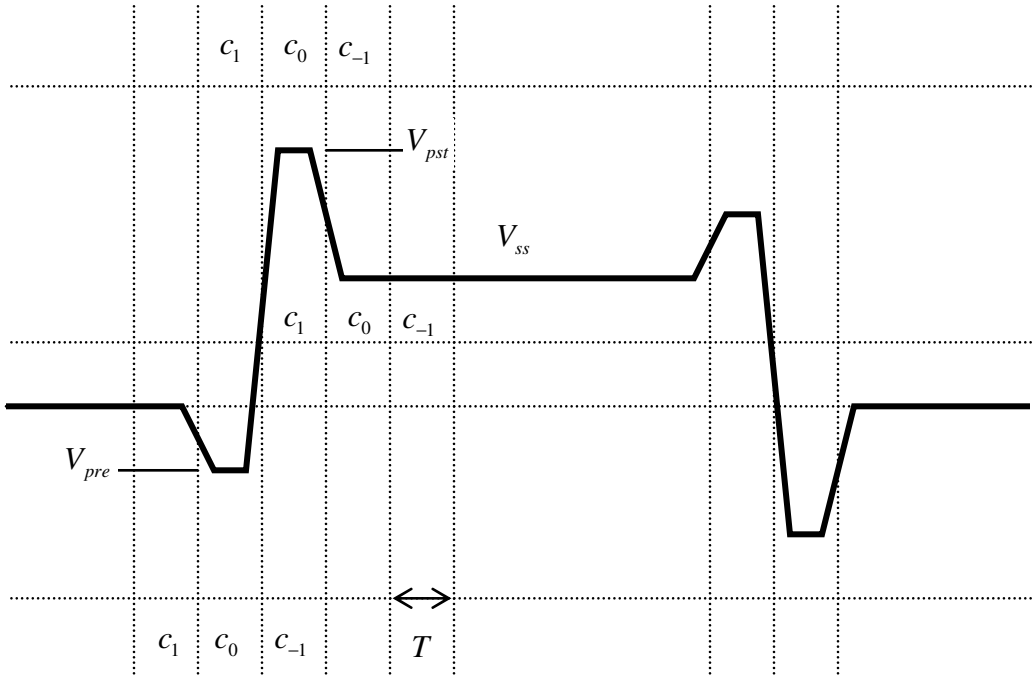
10GBASE-KR Architecture

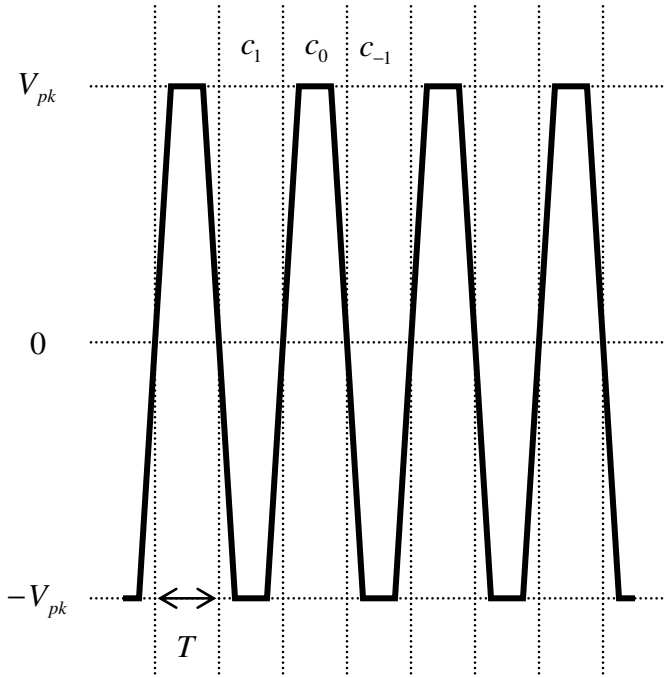
- 10 Gb/s connection between two media access controllers over an electrical backplane
- Low-overhead 64B/66B encoding
- Statistical transition density and DC balance (scrambled)
- Signaling speed is 10.3125 Gbaud
- Optional Forward Error Correction
- Transmitter equalizer programmed by the link partner during start-up
- Adaptive receiver equalization



agere systems

Transmit Equalizer Signal Shaping





$$V_{pre} = -c_1 - c_0 + c_{-1}$$

$$V_{pst} = -c_1 + c_0 + c_{-1}$$

$$V_{ss} = c_1 + c_0 + c_{-1}$$

$$V_{pk} = |c_1| + |c_0| + |c_{-1}|$$

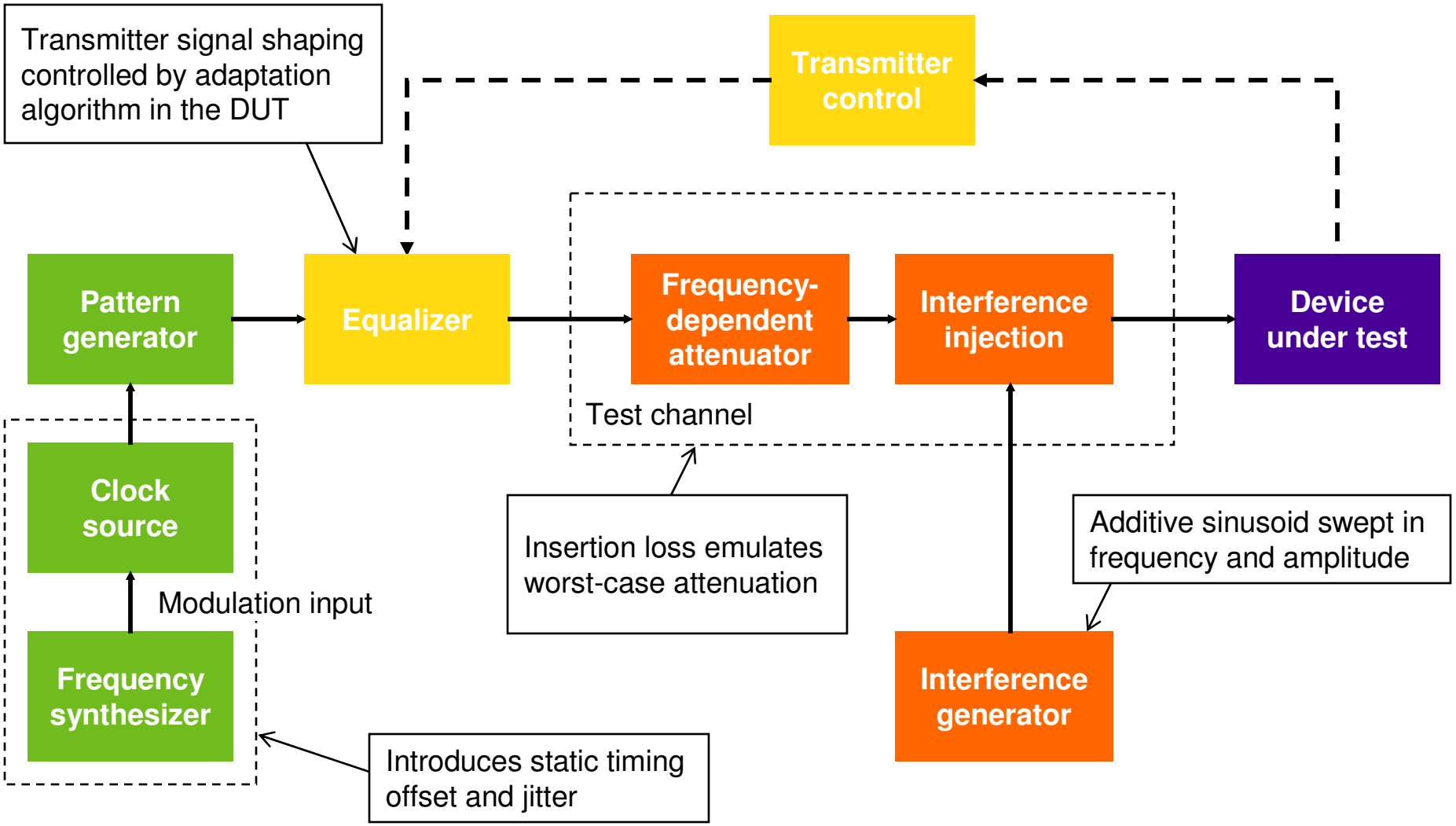
Assumes that the implementation is a three-tap transversal filter with unit delay.

agere systems

Receive Equalizer

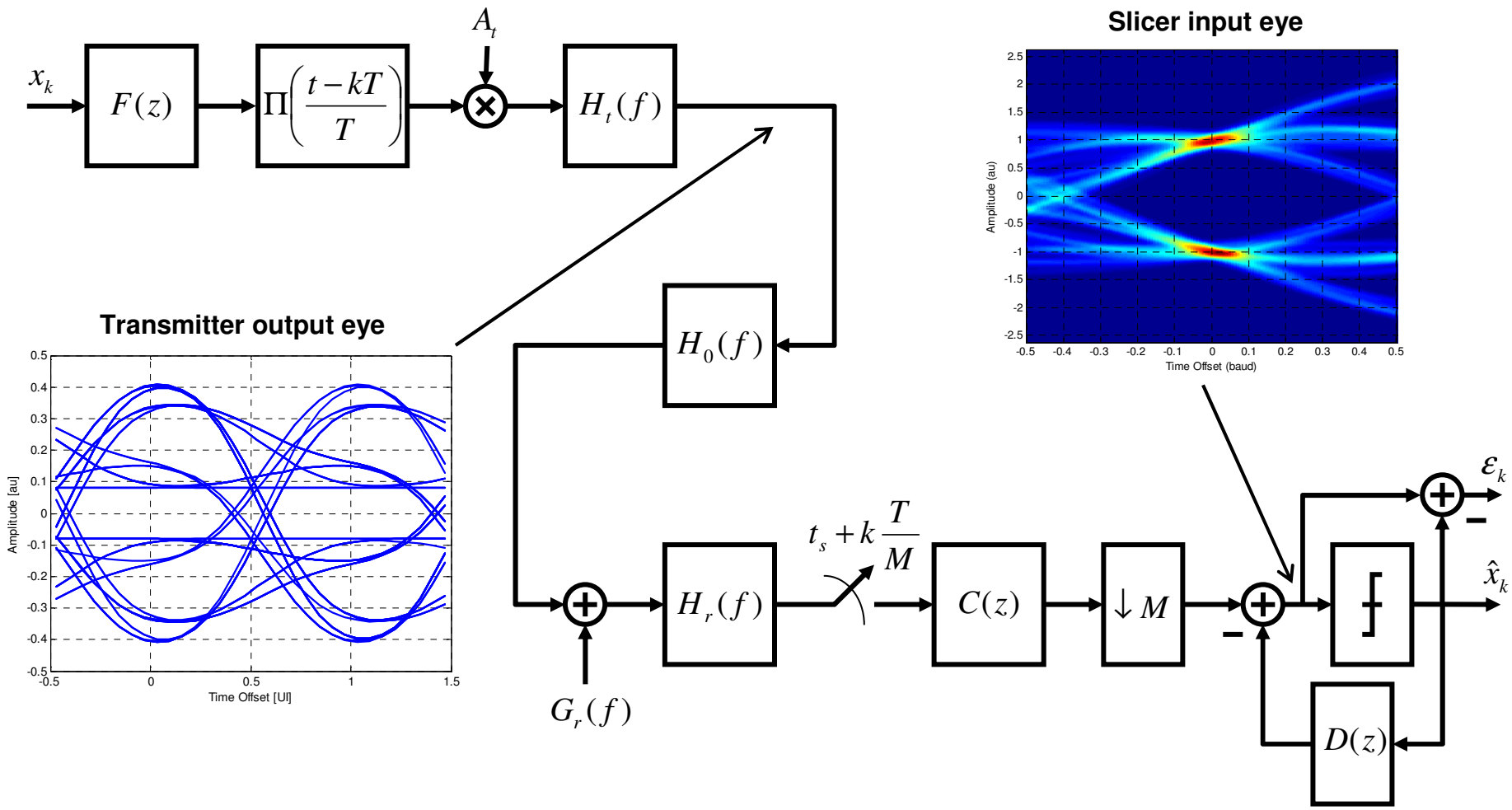
- Requirements developed under the assumption of a 5-tap decision feedback equalizer
- However, the *implementation* of this architecture is not necessarily required for compliance
- Instead, the receiver in question must exhibit an expected level of performance as established via the *interference tolerance* test
- The interference tolerance test examines the ability of the receiver to equalize a high-loss channel in the presence of horizontal (jitter) and vertical (noise) interference

Interference Tolerance Testing



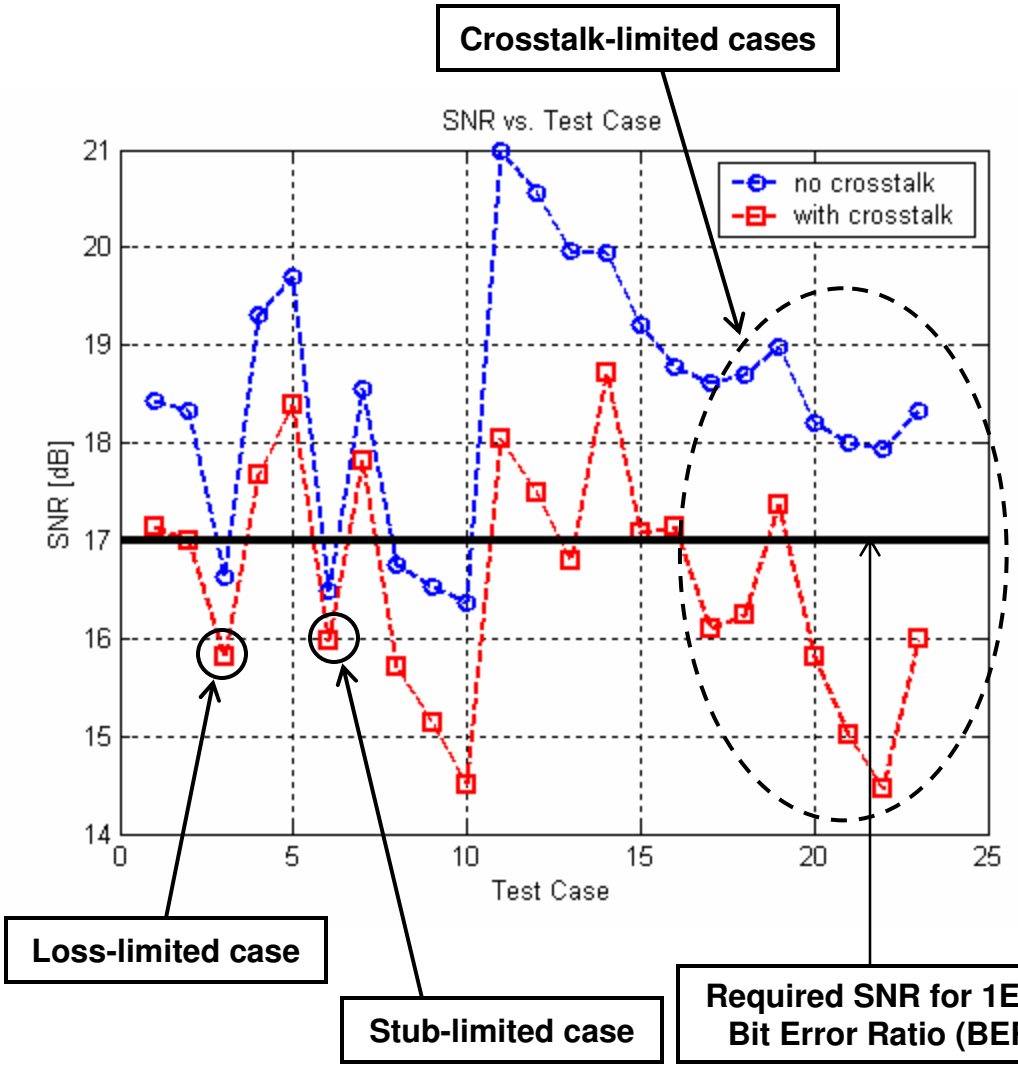
agere systems

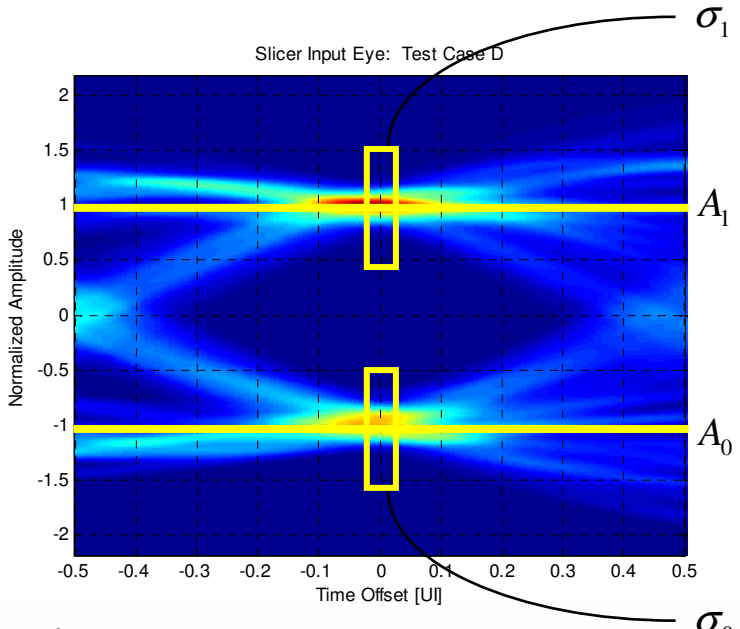
10GBASE-KR Link Model

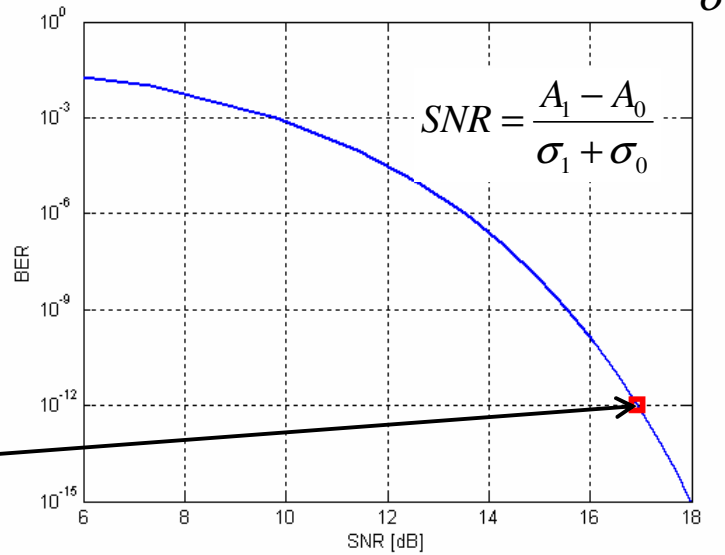


agere systems

10GBASE-KR Performance Simulations





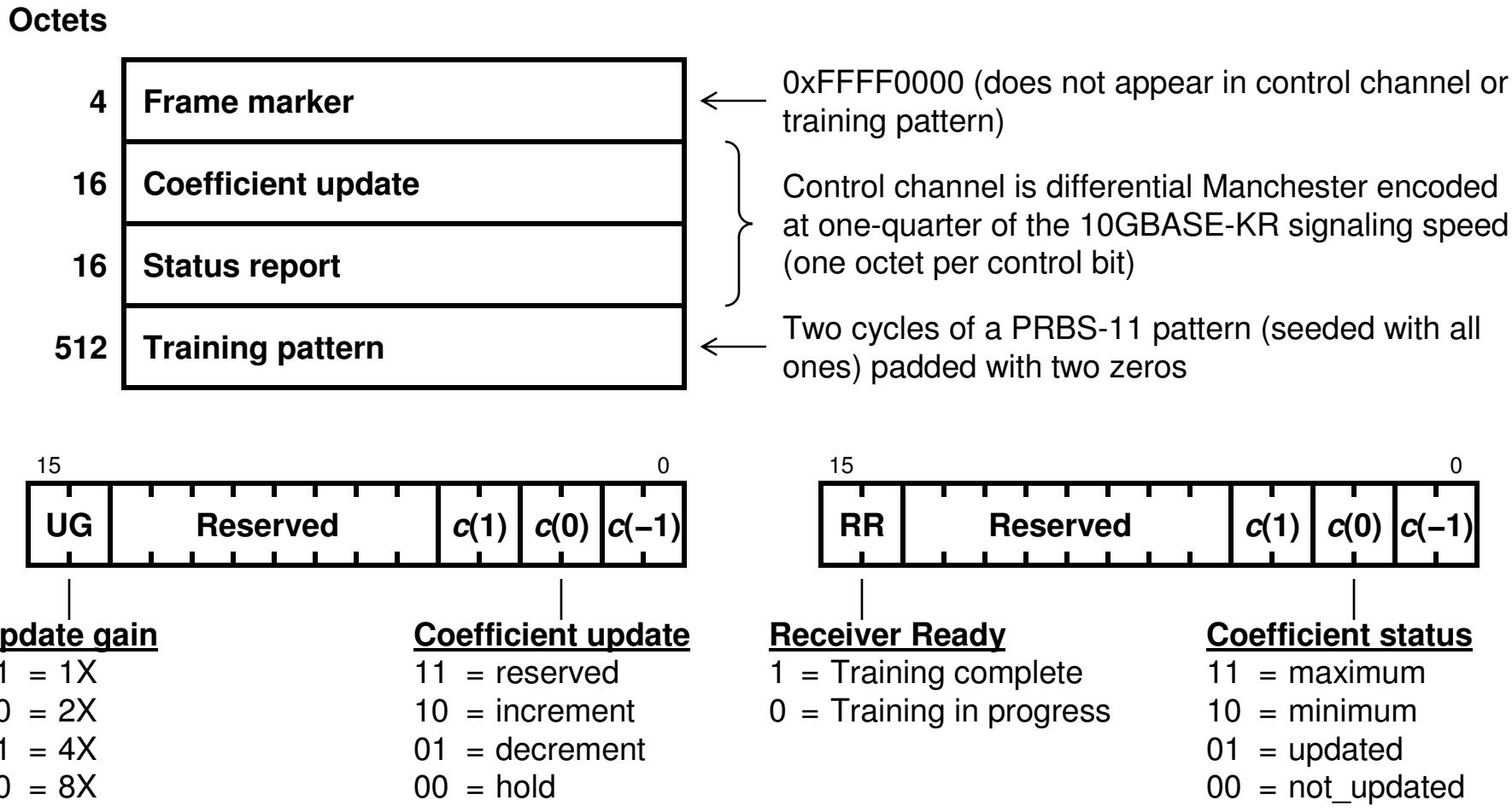


agere systems

10GBASE-KR Start-Up Protocol

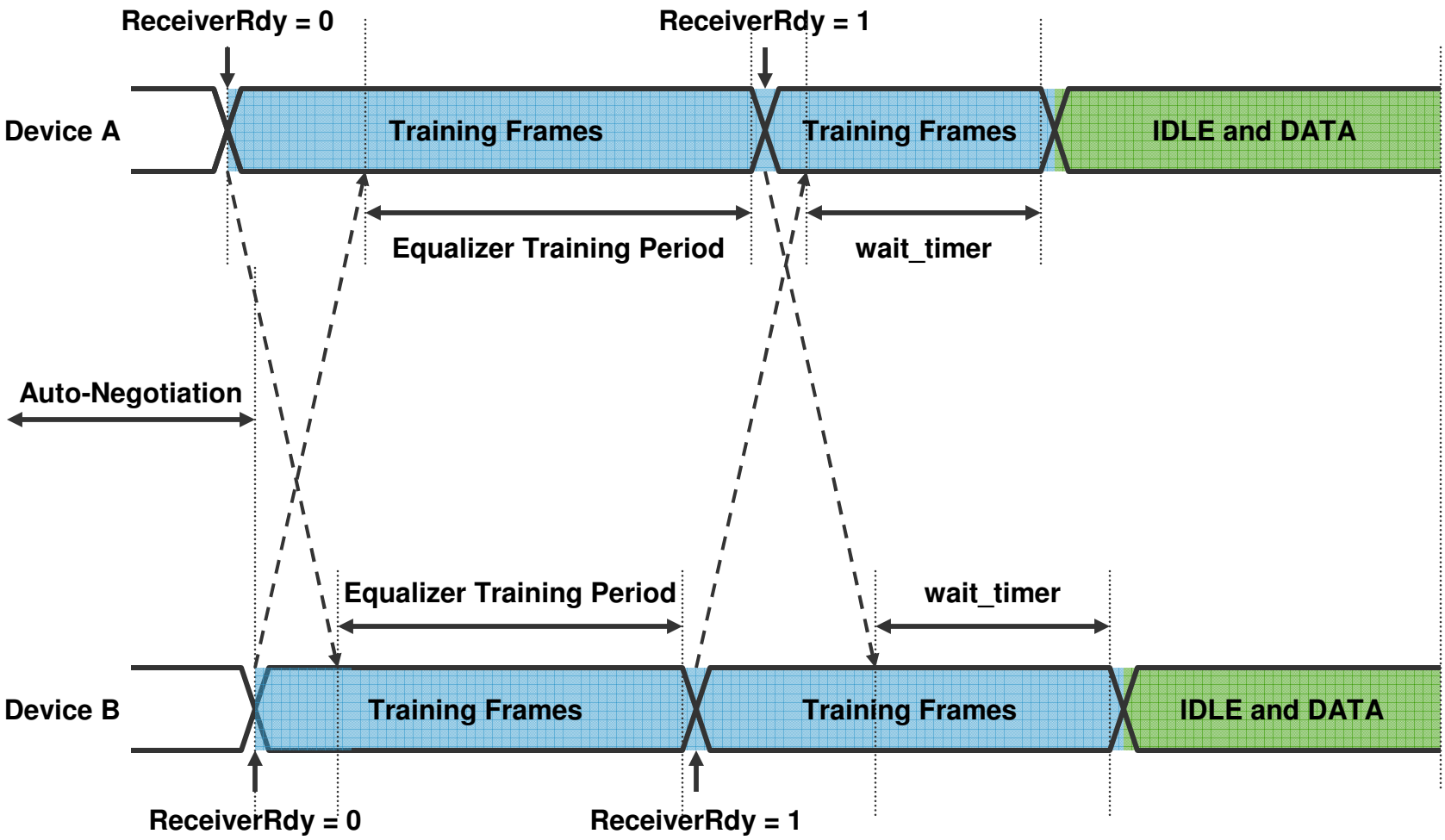
- Optimizes transmitter FIR
- Automatic power control
  - Receiver may steer the transmitter output voltage to the minimum level required for acceptable performance
  - May also mitigate crosstalk
- Optimize receiver equalizer
  - Joint adaptation of transmitter and receiver yields superior solution to independent adaptation
- Accomplished through the exchange of fixed-length training frames

Training Frame Structure



agere systems

Timing Diagram



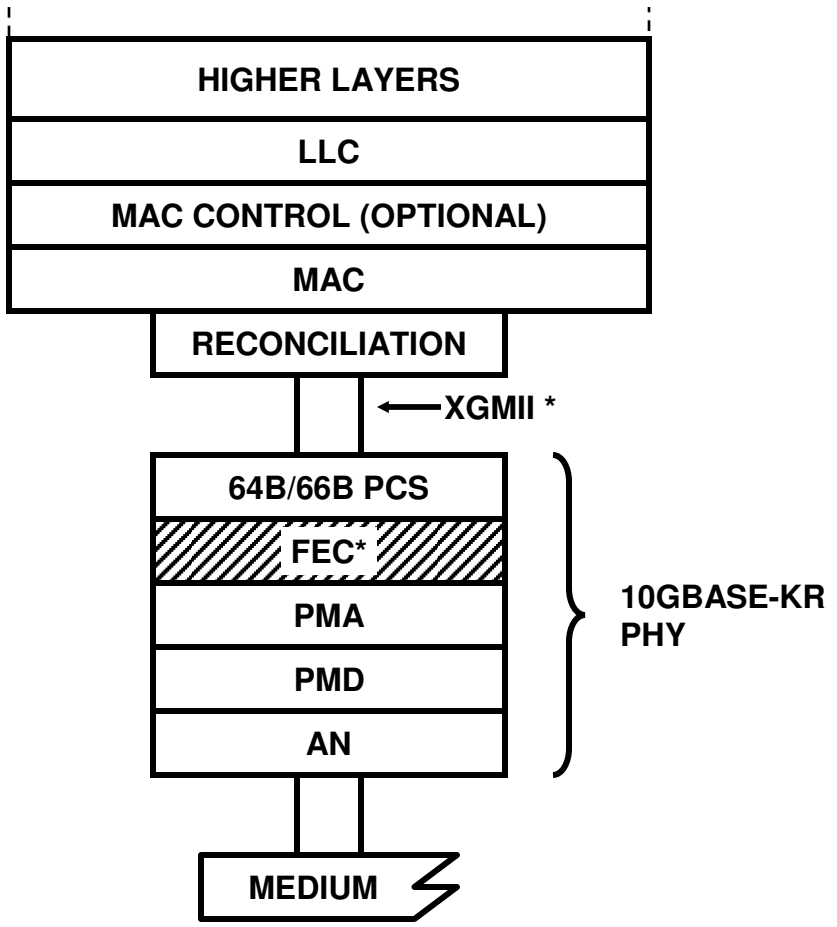
agere systems

Agenda

- Introduction to Backplane Ethernet
- Backplane Architecture and Reference Model
- 10GBASE-KR
- Forward Error Correction for 10GBASE-KR
- Closing Remarks

Forward Error Correction for 10GBASE-KR

- FEC improves 10GBASE-KR link performance at the expense of added complexity and latency
- Not required to meet objectives
- Burst error correction especially suitable for DFE-based receiver architecture
  - Counter-acts error propagation
- Signaling speed is not altered by FEC



agere systems

FEC Frame Structure

- Transcode 66B blocks into FEC frame (32 blocks/frame)
  - T-bit is the second bit of the 66B sync header (0 = control, 1 = data)
- Append 32 parity bits (total frame length is 2112 bits)

T_0	64-bit payload word 0	T_1	64-bit payload word 1	T_2	64-bit payload word 2	T_3	64-bit payload word 3
T_4	64-bit payload word 4	T_5	64-bit payload word 5	T_6	64-bit payload word 6	T_7	64-bit payload word 7
T_8	64-bit payload word 8	T_9	64-bit payload word 9	T_{10}	64-bit payload word 10	T_{11}	64-bit payload word 11
T_{12}	64-bit payload word 12	T_{13}	64-bit payload word 13	T_{14}	64-bit payload word 14	T_{15}	64-bit payload word 15
T_{16}	64-bit payload word 16	T_{17}	64-bit payload word 17	T_{18}	64-bit payload word 18	T_{19}	64-bit payload word 19
T_{20}	64-bit payload word 20	T_{21}	64-bit payload word 21	T_{22}	64-bit payload word 22	T_{23}	64-bit payload word 23
T_{24}	64-bit payload word 24	T_{25}	64-bit payload word 25	T_{26}	64-bit payload word 26	T_{27}	64-bit payload word 27
T_{28}	64-bit payload word 28	T_{29}	64-bit payload word 29	T_{30}	64-bit payload word 30	T_{31}	64-bit payload word 31
32 parity bits							

FEC Encoder and Decoder

- Parity bits computed using a shortened cyclic code (2112, 2080)
- Frame is then scrambled
- Approximately 2 dB coding gain
- Guaranteed error correction for bursts up to 11 bits in length
- Complexity estimated to be 14K gates plus a 64 x 33 dual port RAM at 312.5 MHz
- Latency estimated to be 200 ns

Agenda

- Introduction to Backplane Ethernet
- Backplane Architecture and Reference Model
- 10GBASE-KR
- Forward Error Correction for 10GBASE-KR
- Closing Remarks

Observations

- Backplane channels are not as predictable or easily classified as cabling channels
- Backplane channel length is a misleading metric
  - Significant passband ripple, stub-related resonances, or crosstalk coupling may make short links unworkable
  - With proper material selection and design practices, 1 m links are feasible
- Backplane behavior may differ significantly from what is measured in the lab
  - Ideal terminations are replaced with actual devices
  - Temperature, humidity, and resin tolerances can cause variability in backplanes in the field – margin must be allocated

Conclusions

- IEEE P802.3ap defines a comprehensive Physical Layer suite to support Ethernet as a backplane fabric
  - Addresses the absence of an industry standard for Backplane Ethernet
  - Provides the industry a roadmap to 10 Gb/s serial transmission over an electrical backplane
- Still IEEE P802.3ap is only a piece of an Ethernet fabric solution
  - Improvements to congestion management are necessary to make Ethernet a viable storage or IPC solution
  - Improvements to congestion management enhance networking and transport performance

agere systems

Thank you!