Reference receiver framework for 200G/lane electrical interfaces and PHYs

Adam Healey, Raj Hegde

Broadcom Inc.

IEEE P802.3dj Task Force

January 2024 (r4)

Background

- Recent calculations of Channel Operating Margin (COM) have included a feed-forward equalizer (FFE) and 1-tap decision feedback equalizer (DFE) in the reference receiver
- However, this structure has not been formally adopted
- The method used to optimize the reference receiver equalizer coefficients and sampling time has been a topic of discussion
- It would be beneficial to formally select the reference receiver structure including the method of optimization
- This would provide a solid platform for the evaluation of COM parameters and the selection of parameter values for a baseline proposal
- This presentation proposes a method of optimization that shows benefits over currently employed techniques

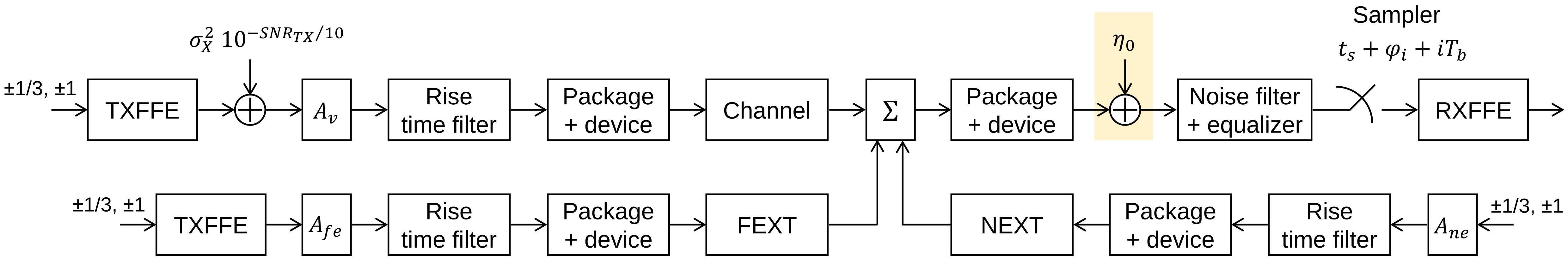
Proposal

- Formalize that the reference receiver is a feed-forward equalizer with a 1-tap decision feedback equalizer
- Optimize equalizer coefficients using the minimum mean-squared error (MMSE) criterion
- This is a well-documented and well-analyzed method of optimization
- See Appendix A for the derivation of equations used in this contribution
- MMSE optimization uses knowledge of the noise at the receiver input
- This requires calculation of the noise autocorrelation function but this is readily done using intermediate results of the existing COM calculation
- Compute a figure of merit (FOM) based on MMSE optimization results
- Choose the sampling time that maximizes this figure of merit (FOM)

Noise autocorrelation function

- The sources of noise considered in the calculation of COM are...
  - Receiver input-referred noise
  - Crosstalk
  - Transmitter output noise
  - Noise resulting from transmitter jitter
- A power spectral density can be defined for each source of noise
- Derive the noise autocorrelation function from the inverse Fourier transform of the sum of the noise power spectral densities

Receiver noise power spectral density

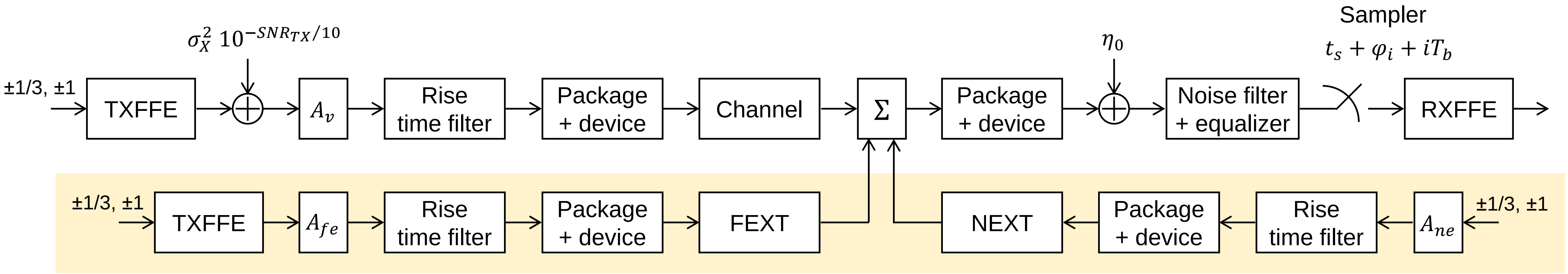


Given the transfer function of the receiver noise filter $H_r(f)$, the transfer function of the continuous-time equalizer $H_{ctf}(f)$, and one-sided power spectral density η_0 (converted to units V^2/Hz) ...

$\tilde{S}_{rn}(f) = \frac{\eta_0}{2}\left|H_r(f)H_{ctf}(f)\right|^2$ is the receiver power spectral density at sampler input

$S_{rn}(\theta) = \sum_m \tilde{S}_{rn}\left(\frac{\theta + 2\pi m}{2\pi T_b}\right) \quad -\pi < \theta \leq \pi$ is the folded power spectral density corresponding to the sampled noise at the RXFFE input where $T_b = 1/f_b$ is the unit interval

Crosstalk power spectral density



Given the signal power $\sigma_X^2 = \dfrac{L^2-1}{3(L-1)^2}$ where L is the number of signal levels and ...

... given the pulse response for k^{th} crosstalk aggressor $h^{(k)}(t)$

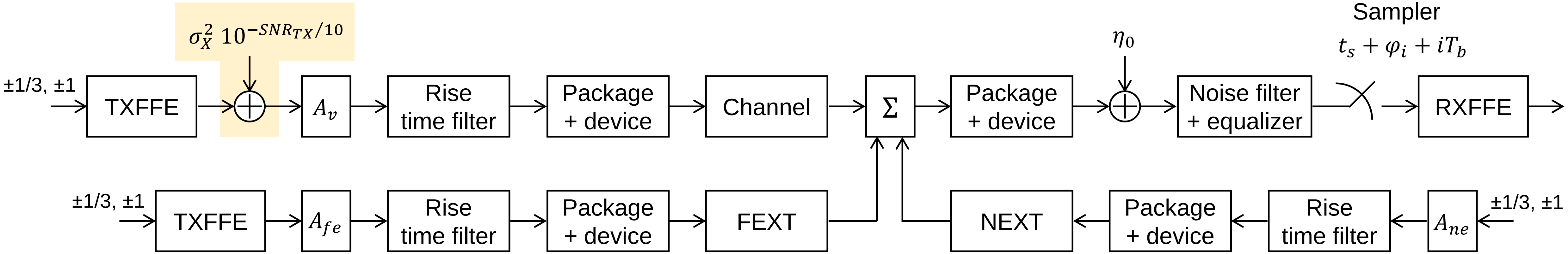
$$h_{xn}^{(k)}(i) = h^{(k)}\left(\frac{m+iM}{Mf_b}\right)$$ where M is the number of samples per unit interval

Note that m is chosen to maximize

$$\sum_i \left[h_{xn}^{(k)}(i)\right]^2$$

$$S_{xn}^{(k)}(\theta) = \sigma_X^2 \left|\mathcal{F}\left\{h_{xn}^{(k)}(i)\right\}\right|^2 / f_b$$ is the power spectral density of the k^{th} crosstalk aggressor at the RXFFE input where $\mathcal{F}\{x\}$ is the Fourier Transform of x

Transmitter noise power spectral density



Given the transmitter signal-to-noise ratio SNR_{TX}, the transfer function of the rise time filter $H_t(f)$, and the voltage transfer function of the victim signal path $H_{21}^{(0)}(f)$, ...

$$\left.\begin{aligned} H_{tn}(f) &= H_t(f)H_{21}^{(0)}(f)H_r(f)H_{ctf}(f) \\ \tilde{h}_{tn}(t) &= \mathcal{F}^{-1}\{A_v T_b \operatorname{sinc}(fT_b)H_{tn}(f)\} \end{aligned}\right\}$$

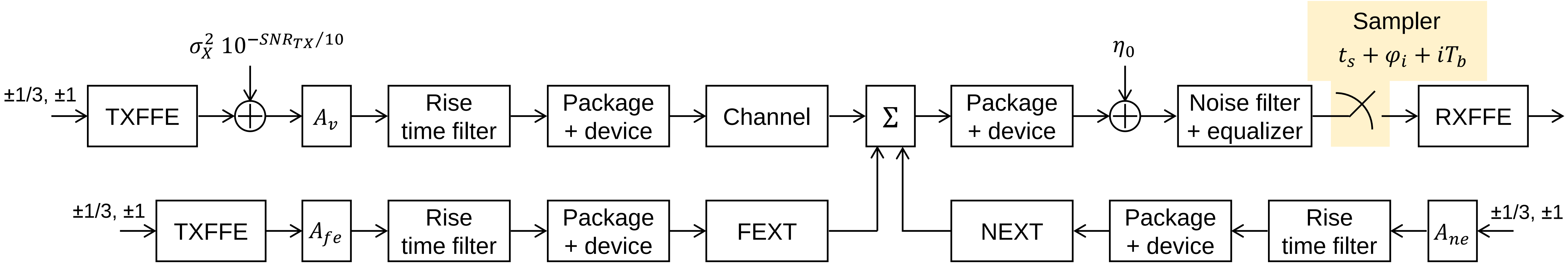
is the pulse response of the victim signal path <u>excluding the TXFFE response</u>

$$h_{tn}(i) = \tilde{h}_{tn}(t_s + iT_b)$$

$$S_{tn}(\theta) = \sigma_X^2 10^{-SNR_{TX}/10}\, |\mathcal{F}\{h_{tn}(i)\}|^2/f_b$$

is the power spectral density of the transmitter noise at the RXFFE input

Power spectral density of noise due to jitter

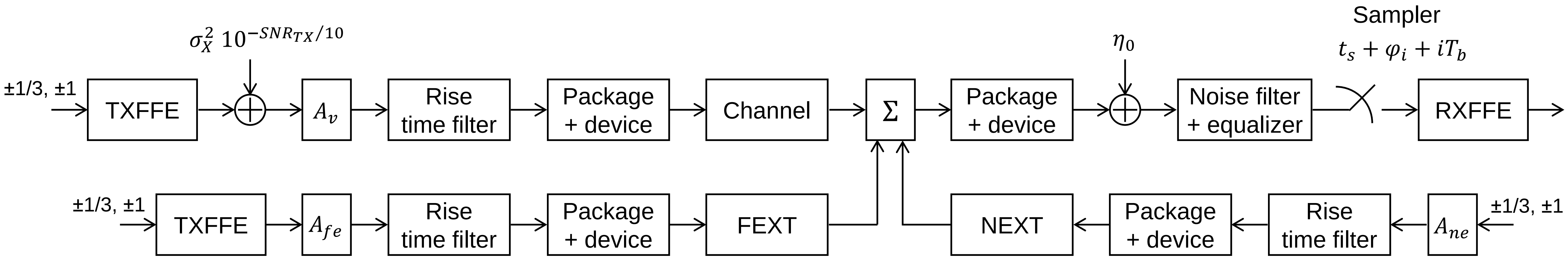


Given the peak Dual-Dirac jitter A_{DD}, the RMS random jitter σ_{RJ}, and $h_J(i)$ which is the slope of victim signal path pulse response around the sampled values ...

$$S_{jn}(\theta) = \sigma_X^2(A_{DD}^2 + \sigma_X^2)\left|\mathcal{F}\{h_J(i)\}\right|^2/f_b$$

is the power spectral density of the noise due to jitter as observed at the RXFFE input

Noise autocorrelation function definition



$S_n(\theta) = S_{rn}(\theta) + \sum_{k=1}^{K-1} S_{xn}^{(k)}(\theta) + S_{tn}(\theta) + S_{jn}(\theta)$ is the power spectral density of the total (sampled) noise at the RXFFE input

$R_n(i) = \mathcal{F}^{-1}\{S_n(\theta)\}\, f_b$ is the noise autocorrelation at the RXFFE input

$\boldsymbol{R}_{nn}$ is the noise autocorrelation matrix which is a diagonal-constant (Toeplitz) matrix whose first row and column are $R_n(i)$ for i = 0 to $N_w - 1$ where N_w is the number of feed-forward filter taps

Proposed coefficient optimization procedure

- Given the pulse response for the victim signal path $h^{(0)}(t)$ and the sampling time t_s, ...
- Define $\boldsymbol{h}$ to be the vector $[h_{-d_h}$, ..., h_0, ..., $h_{N-d_h-1}]$ where h_m is $h^{(0)}(t_s + mT)$
- Define the delay d to be $d_h + d_w$ where d_w is the feed-forward filter delay (equal to the number of pre-cursor taps)
- Define $\boldsymbol{H}$ to be a constant-diagonal (Toeplitz) matrix whose first column is $\boldsymbol{h}$ followed by $N_w - 1$ zeros whose first row is h_{-d_h} followed by $N_w - 1$ zeros
- Define $\boldsymbol{h}_0$ to be row $d + 1$ from $\boldsymbol{H}$
- Define $\boldsymbol{H}_b$ to be rows $d + 2$ to $d + N_b + 1$ from $\boldsymbol{H}$ (N_b is the feedback filter length)
- Define $\boldsymbol{R}$ to be $\boldsymbol{H}^T\boldsymbol{H} + \boldsymbol{R}_{nn}/\sigma_X^2$ where $\boldsymbol{R}_{nn}$ is the previously defined noise autocorrelation matrix, σ_X^2 is the signal power, and a T exponent denotes the matrix transpose
- Solve the following matrix equation (ignoring the resulting value of λ)

$$\begin{bmatrix} \boldsymbol{w} \\ \boldsymbol{b} \\ \lambda \end{bmatrix} = \begin{bmatrix} \boldsymbol{R} & -\boldsymbol{H}_b^T & -\boldsymbol{h}_0^T \\ -\boldsymbol{H}_b & \boldsymbol{I}_b & \boldsymbol{z}_b^T \\ \boldsymbol{h}_0 & \boldsymbol{z}_b & 0 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{h}_0^T \\ \boldsymbol{z}_b^T \\ 1 \end{bmatrix}$$

where $\boldsymbol{I}_b$ is the N_b x N_b identity matrix and $\boldsymbol{z}_b$ is a row vector of N_b zeros

Proposed coefficient optimization procedure, continued

- Apply specified minimum and maximum limits to $\boldsymbol{b}$ to yield $\boldsymbol{b}_{lim}$
- If $\boldsymbol{b}_{lim}$ is not equal to $\boldsymbol{b}$, then solve the following matrix equation to update $\boldsymbol{w}$

$$\begin{bmatrix} \boldsymbol{w} \\ \lambda \end{bmatrix} = \begin{bmatrix} \boldsymbol{R} & -\boldsymbol{h}_0^T \\ \boldsymbol{h}_0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{h}_0^T + \boldsymbol{H}_b^T \boldsymbol{b}_{lim} \\ 1 \end{bmatrix}$$

(again ignoring λ)

- Apply specified minimum and maximum limits to $\boldsymbol{w}$ to yield $\boldsymbol{w}_{lim}$
- If $\boldsymbol{w}_{lim}$ is not equal to $\boldsymbol{w}$, then ...
  - Normalize $\boldsymbol{w}_{lim}$ by $\boldsymbol{h}_0 \boldsymbol{w}_{lim}$ so that the amplitude of the equalized pulse is 1
  - Update $\boldsymbol{b} = \boldsymbol{H}_b \boldsymbol{w}_{lim}$
  - Apply specified minimum and maximum limits to $\boldsymbol{b}$ to yield $\boldsymbol{b}_{lim}$
- Compute the mean-squared error

$$\sigma_e^2 = \sigma_X^2 \left(\boldsymbol{w}_{lim}^T \boldsymbol{R}\, \boldsymbol{w}_{lim} + 1 + \boldsymbol{b}_{lim}^T \boldsymbol{b}_{lim} - 2\boldsymbol{w}_{lim}^T \boldsymbol{h}_0^T - 2\boldsymbol{w}_{lim}^T \boldsymbol{H}_b^T \boldsymbol{b}_{lim} \right)$$

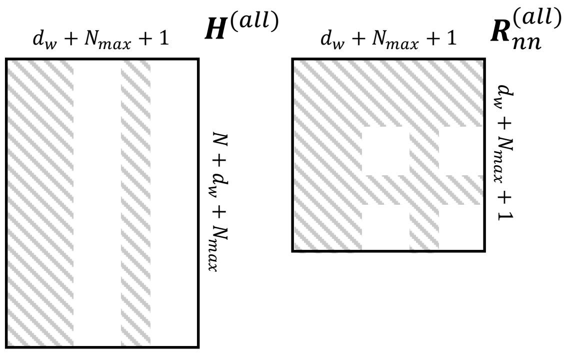
- Compute the figure of merit (FOM)

$$\mathrm{FOM} = 20\log_{10}\left(\frac{R_{LM}/(L-1)}{\sigma_e} \right)$$

where R_{LM} is the specified level separation mismatch ratio (note that the amplitude of the equalized pulse is 1)

Proposed floating tap optimization procedure

- Given N_{fix} fixed-position taps, d_w pre-cursor taps, N_g banks of floating taps, N_f taps per bank, and highest allowed floating tap position N_{max}
- Define $\boldsymbol{H}^{(all)}$ to be a constant-diagonal matrix whose first column is $\boldsymbol{h}$ followed by $d_w + N_{max}$ zeros and whose first row is h_{-d_h} followed by $d_w + N_{max}$ zeros
- Define $\boldsymbol{R}_{nn}^{(all)}$ to be a diagonal-constant matrix whose first column and row are $R_n(i)$ for i = 0 to $d_w + N_{\max}$
- Define $\boldsymbol{i}$ to a vector of N_w tap indices consisting of indices 1 to N_{fix} for fixed-position taps and $N_f \times N_g$ additional indices for floating taps
- Define $\boldsymbol{H}$ to be the columns $\boldsymbol{i}$ from $\boldsymbol{H}^{(all)}$ and $\boldsymbol{R}_{nn}$ to be columns $\boldsymbol{i}$ and rows $\boldsymbol{i}$ from $\boldsymbol{R}_{nn}^{(all)}$



Shaded areas are the sub-matrices defined by $\boldsymbol{i}$ (yielding $\boldsymbol{H}$ and $\boldsymbol{R}_{nn}$)

- Solve for the equalizer coefficients and calculate the FOM using the previously defined algorithm
- Floating tap indices in $\boldsymbol{i}$ are chosen to maximize the FOM
- Floating tap indices within a bank must be contiguous and must not overlap with the indices of other floating tap banks or fixed-position taps
- Any search method that yields the correct answer (maximum FOM) should be allowed

Proposed sampling time optimization procedure

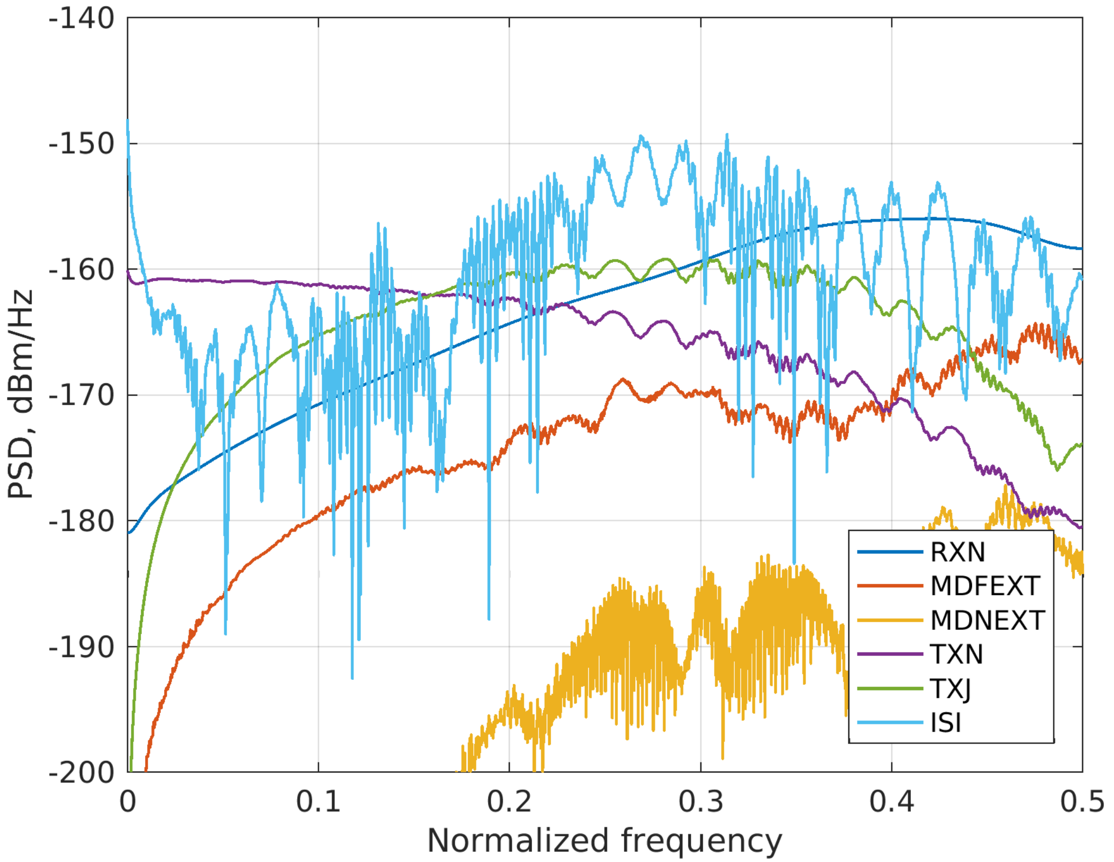
- Find the value of t_s that maximizes the FOM
- Any search method that yields the correct answer (maximum FOM) should be allowed
- Constraints on the search range and allowances for t_s granularity can be considered

Additional considerations

- Given the FOM-optimized sampling phase and equalizer coefficients, it is straightforward apply the feed-forward equalizer transfer function to the time-domain responses and calculate the probability density functions for the noise and interference at the equalizer output
- If needed, these steps can be defined in detail in a future contribution

Residual inter-symbol interference (ISI)

- An example of the power spectral density for each impairment at the equalizer output is shown
- The power spectral density of residual ISI, $S_{isi}(\theta)$, is included
- It may be computed using the following equations
- It may be useful for the calculation of performance improvements from maximum likelihood sequence estimation



$$h_{isi}(i) = \begin{cases} \boldsymbol{h}_i \boldsymbol{w} - 1 & i = d + 1 \\ \boldsymbol{h}_i \boldsymbol{w} - b_{i-d-1} & 1 \le i - d - 1 \le N_b \\ \boldsymbol{h}_i \boldsymbol{w} & \text{otherwise} \end{cases}$$

where $\boldsymbol{h}_i$ is row i from $\boldsymbol{H}$

$$S_{isi}(\theta) = \sigma_X^2 \, |\mathcal{F}\{h_{isi}(i)\}|^2 / f_b$$

Impact of the proposed optimization method

- Channel Operating Margin (COM) is computed using MMSE optimization
- Results are compared to COM computed using the algorithm described in mellitz_3dj_elec_01_230831 (hereafter referred to as the *force* algorithm)
- The same configuration is used to generate both sets of results for apples-to-apples comparisons

Test case definition

KR channel source files	Number of cases
shanbhag_3dj_02_2305	4
weaver_3dj_02_2305	36
weaver_3dj_elec_01_230622	4
mellitz_3dj_02_elec_230504	27
mellitz_3dj_03_elec_230504	25
akinwale_3dj_01_2310	7
Total	103

CR channel source files	Number of cases
shanbhag_3dj_01_2305	6
kocsis_3dj_02_2305	5
lim_3dj_03_230629	1
lim_3dj_04_230629	1
lim_3dj_07_2309	1
akinwale_3dj_02_2311	4
weaver_3dj_02_2311	12
Total	30

Package class A

Parameter	Setting	Units	Information
package_tl_gamma0_a1_a2	[5e-4 8.9e-4 2e-4]		
package_tl_tau	0.006141	ns/mm	
package_Z_c	[87.5 87.5 ; 92.5 92.5]	Ohm	
z_p select	1		
z_p (TX)	[34 ; 1.8]	mm	[test cases]
z_p (NEXT)	[34 ; 1.8]	mm	[test cases]
z_p (FEXT)	[34 ; 1.8]	mm	[test cases]
z_p (RX)	[32 ; 1.8]	mm	[test cases]
C_p	[0.4e-4 0.4e-4]	nF	[TX RX]

Package class B

Parameter	Setting	Units	Information
package_tl_gamma0_a1_a2	[5e-4 6.5e-4 3e-4]		
package_tl_tau	0.006141	ns/mm	
package_Z_c	[92 92 ; 70 70 ; 80 80 ; 100 100]	Ohm	
z_p select	1		
z_p (TX)	[46 ; 1 ; 1 ; 0.05]	mm	[test cases]
z_p (NEXT)	[46 ; 1 ; 1 ; 0.05]	mm	[test cases]
z_p (FEXT)	[46 ; 1 ; 1 ; 0.05]	mm	[test cases]
z_p (RX)	[44 ; 1 ; 1 ; 0.05]	mm	[test cases]
C_p	[0.4e-4 0.4e-4]	nF	[TX RX]

133 channels x 2 package classes = 266 test cases

COM configuration used for testing (not a baseline proposal)

Parameter	Setting	Units	Information
f_b	106.25	GBd	
f_min	0.05	GHz	
Delta_f	0.01	GHz	
C_d	[0.4e-4 0.9e-4 1.1e-4 ; 0.4e-4 0.9e-4 1.1e-4]	nF	[TX ; RX]
L_s	[0.13 0.15 0.14 ; 0.13 0.15 0.14]	nH	[TX ; RX]
C_b	[0.3e-4 0.3e-4]	nF	[TX RX]
R_0	50	Ohm	
R_d	[50 50]	Ohm	[TX RX]
A_v	0.413	V	
A_fe	0.413	V	
A_ne	0.45	V	
L	4		
M	32		
f_r	0.58	*fb	
c(0)	1		min
c(-1)	0		[min:step:max]
c(-2)	0		[min:step:max]
c(-3)	0		[min:step:max]
c(-4)	0		[min:step:max]
c(1)	0		[min:step:max]
N_b	1		
b_max(1)	0.85		
b_max(2..N_b)	0		
b_min(1)	0		
b_min(2..N_b)	0		
g_DC	0	dB	[min:step:max]
f_z	1e100	GHz	
f_p1	1e100	GHz	
f_p2	1e100	GHz	
g_DC_HP	[-5:0.5:0]	dB	[min:step:max]
f_HP_PZ	1.328125	GHz	

No TX FFE

1-tap DFE

Low-frequency equalizer only

Parameter	Setting	Units	Information
DER_0	2e-4		
T_r	0.004	ns	
FORCE_TR	1	logical	
PMD_type	C2C		
TDR	0	logical	
ERL	0	logical	
EW	0	logical	
MLSE	0	logical	
ts_anchor	1		1 for pulse peak
sample_adjustment	[-16 16]		
Local Search	2		
sigma_RJ	0.01	UI	
A_DD	0.02	UI	
eta_0	6e-09	V^2/GHz	
SNR_TX	33	dB	
R_LM	0.95		

No MLSE

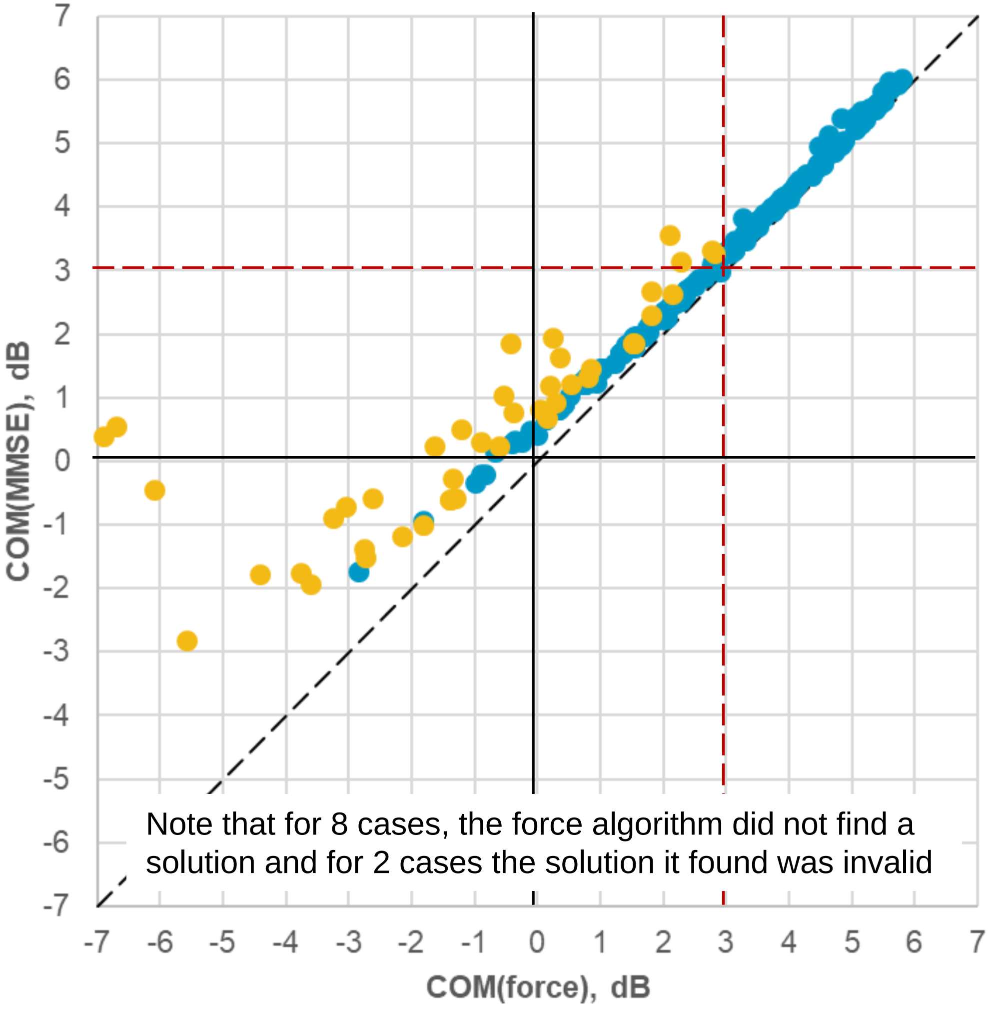
Sample time search

Parameter	Setting	Information
ffe_pre_tap_len	5	
ffe_post_tap_len	10	
ffe_tap_step_size	0	
ffe_main_cursor_min	1	
ffe_pre_tap1_max	1	
ffe_post_tap1_max	1	
ffe_tapn_max	1	
N_g	0 or 1	Number of floating tap groups
N_f	4	Taps per group
N_max	60	Maximum floating tap index

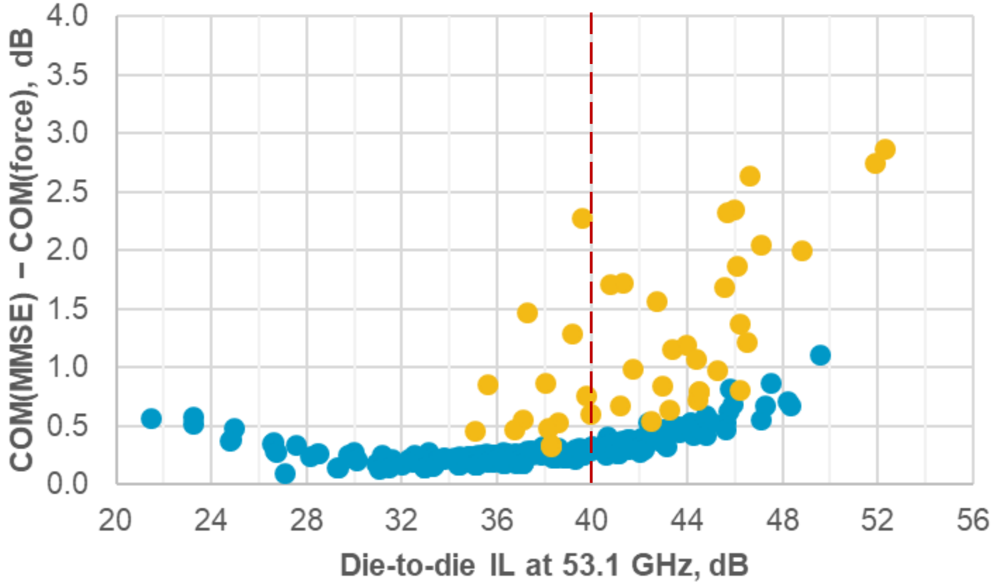
16-tap FFE with constraints and 0 or 1 banks of 4 floating taps

MLSE = maximum likelihood sequence estimation

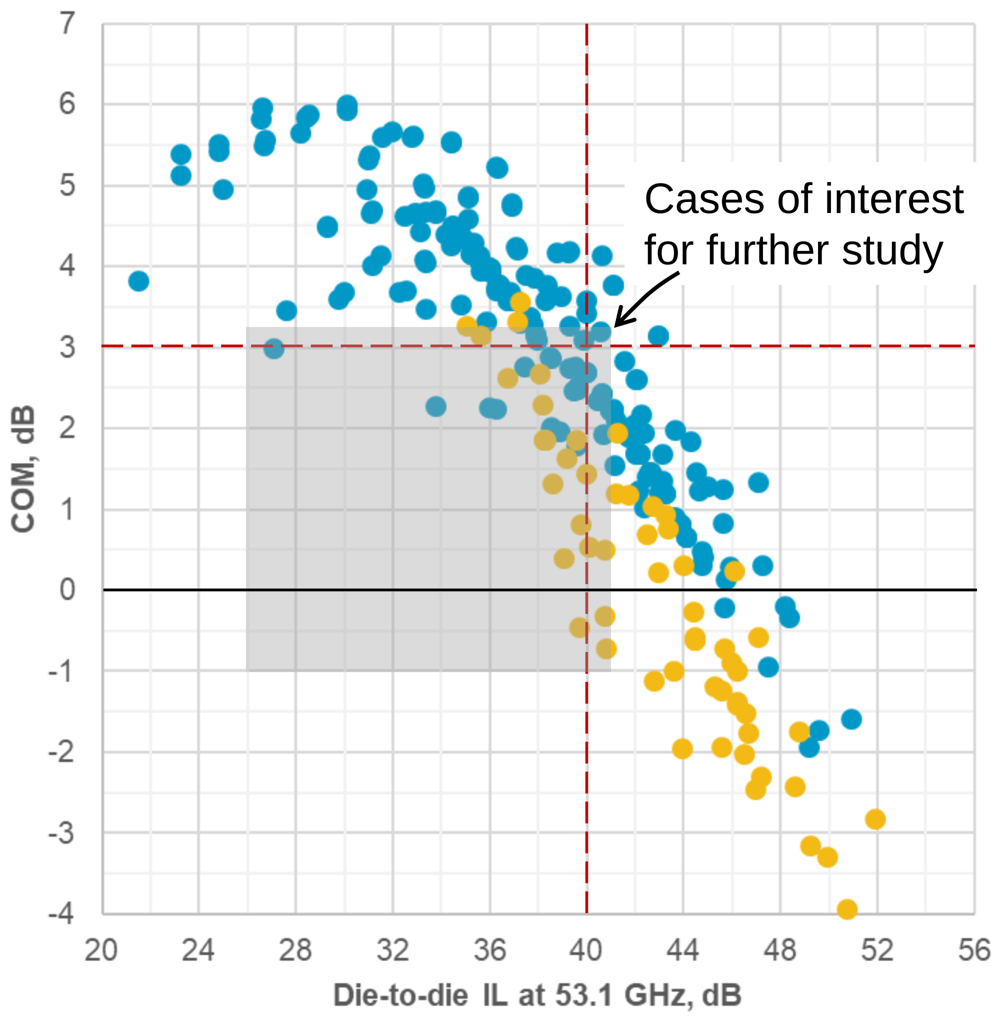
Comparison of results without floating taps



- KR test cases are indicated by **blue** dots
- CR test cases are indicated by **gold** dots
- MMSE optimization consistently yields better results than the force algorithm



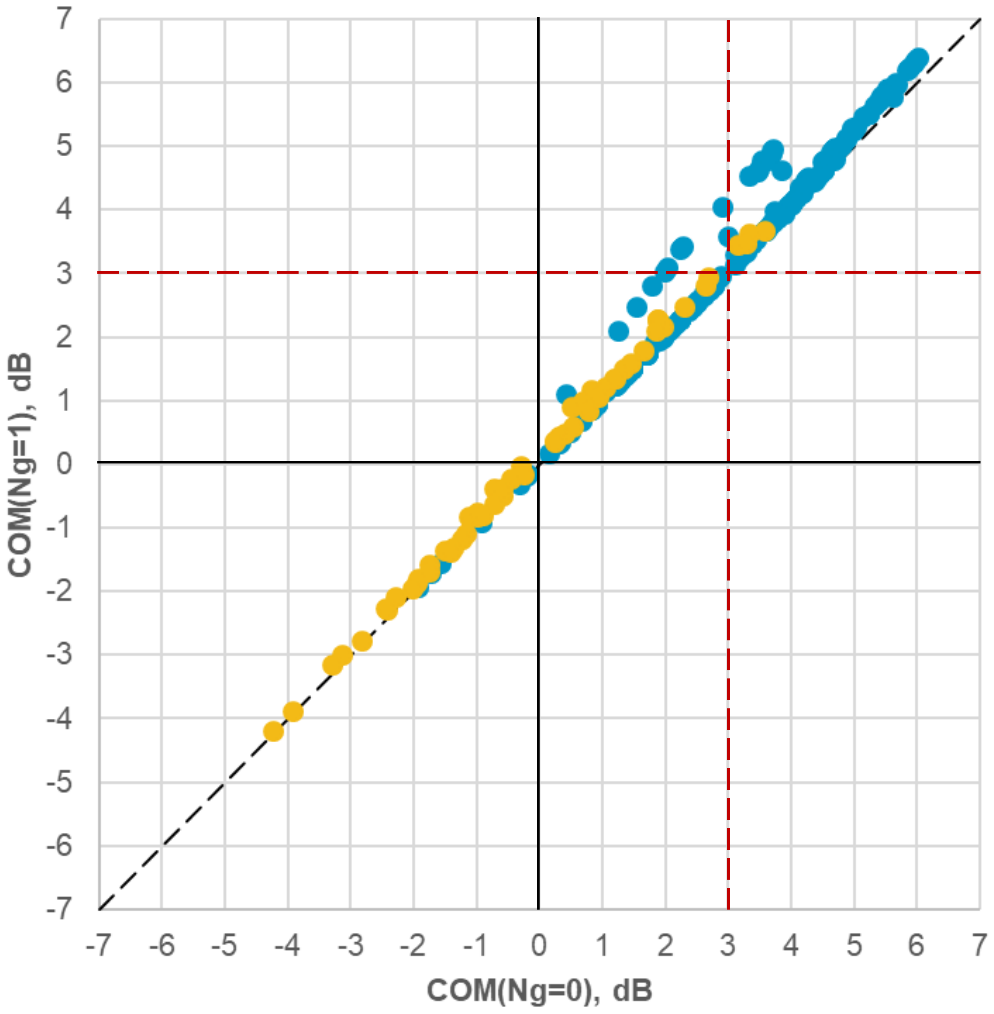
Results using MMSE optimization without floating taps



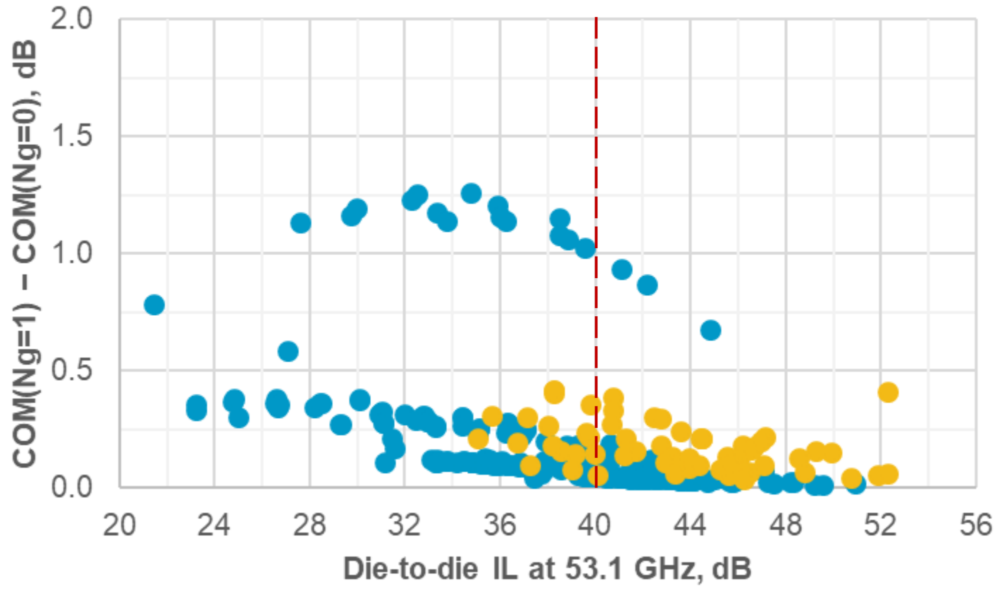
- KR test cases are indicated by **blue** dots
- CR test cases are indicated by **gold** dots
- Improved optimization yields encouraging results despite a lower reference receiver complexity (N_w = 16)
- Results do not include MLSE

- Package models used in this study may not be consistent with host loss assumed for certain CR channels

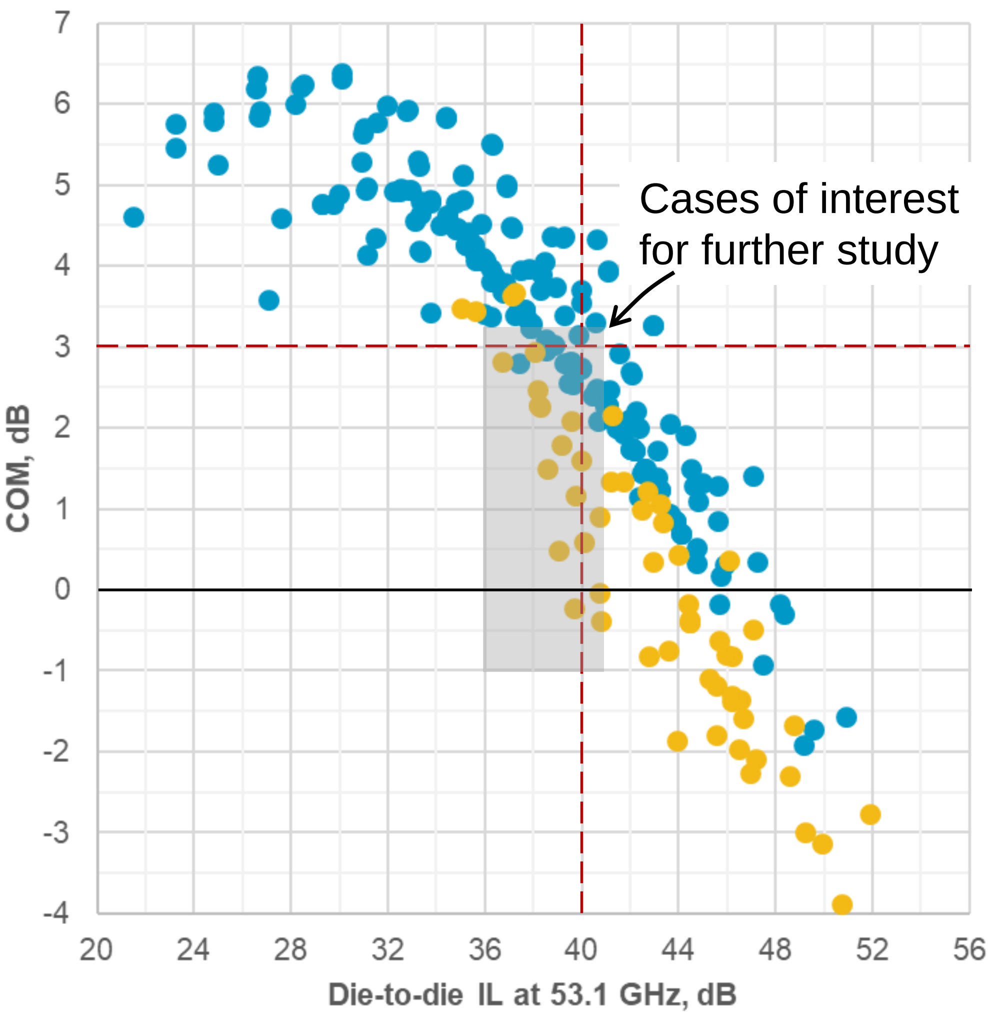
Impact of floating taps



- KR test cases are indicated by **blue** dots
- CR test cases are indicated by **gold** dots
- Floating taps can improve COM via the mitigation of reflections



Results using MMSE optimization with floating taps



- KR test cases are indicated by **blue** dots
- CR test cases are indicated by **gold** dots
- Results do not include MLSE
- Improvements for the cases in the shaded area may be identified via further analysis

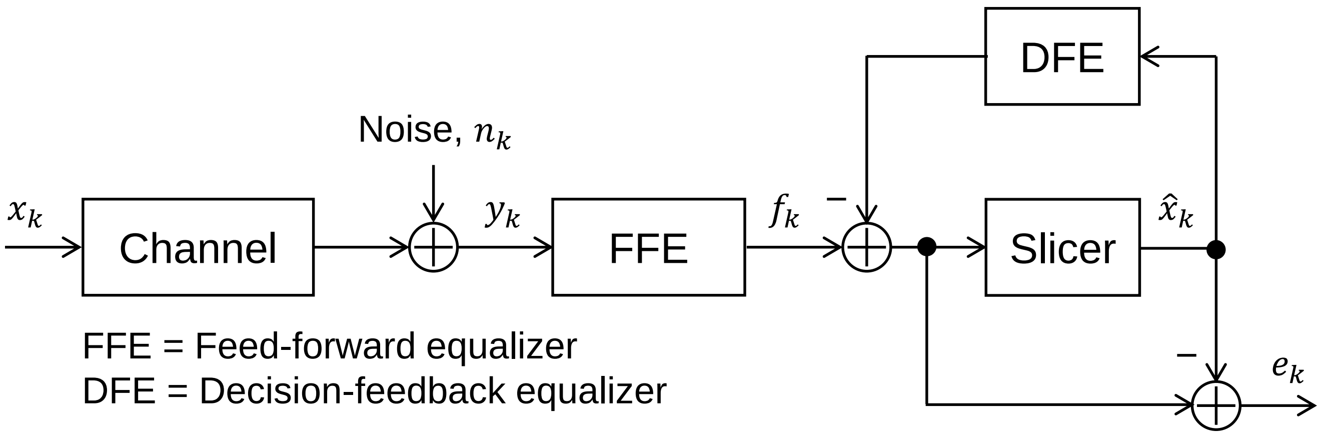
Summary and conclusions

- MMSE optimization is a textbook approach to the determination of FFE and DFE coefficients
- It provides better COM for a given reference receiver than optimization techniques currently in use
- It offers opportunities to reduce the complexity of the reference receiver and / or increase allowances for impairments
- Noise autocorrelation function used for MMSE optimization can then be leveraged to calculate expected performance improvement from MLSE
- Floating feed-forward taps can be included if necessary
- The optimal choice of sampling phase is the one that maximizes FOM
- Adoption of a reference receiver framework is an important step toward baseline proposals

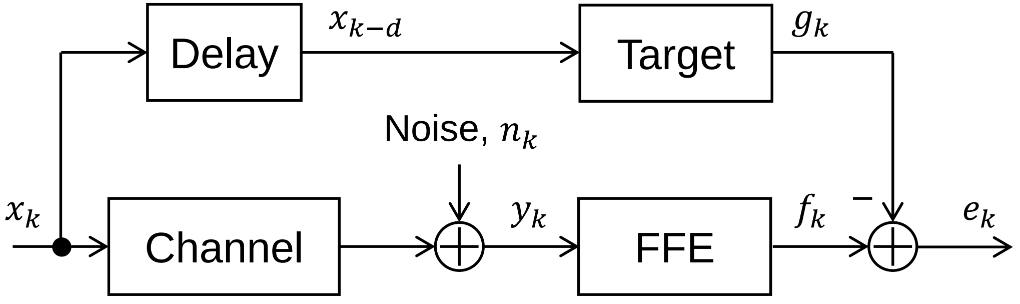
Appendix A

Derivation of minimum mean-squared error equalizer coefficients

System model



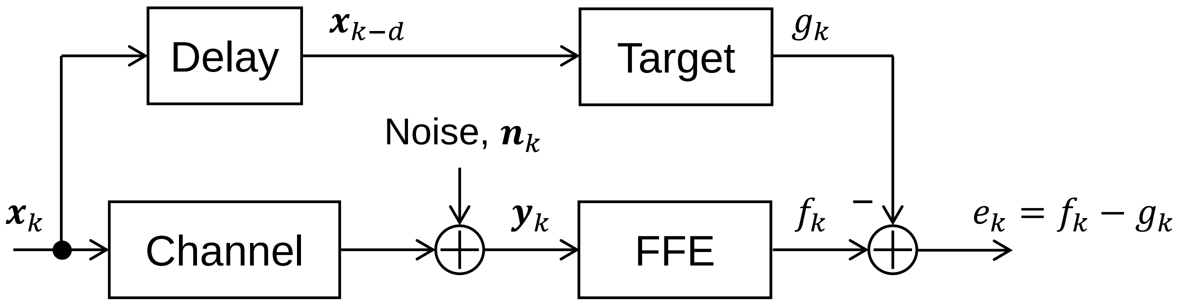
FFE = Feed-forward equalizer
DFE = Decision-feedback equalizer



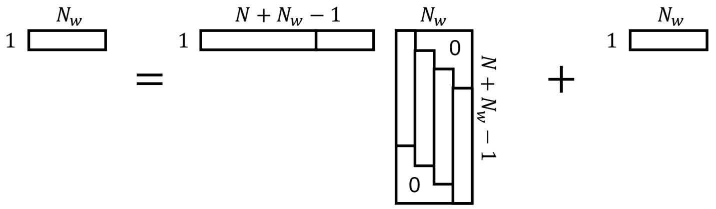
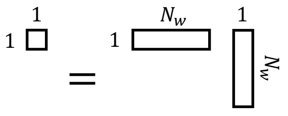
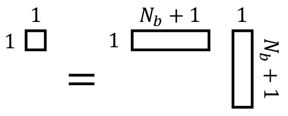
Equivalent model when slicer outputs are correct i.e., $\hat{x}_k = x_{k-d}$

Channel output	FFE output	Target signal
$y_k = \sum_{m=-d_h}^{N-d_h-1} h_m\, x_{k-m} + n_k$ h_m is a coefficient of the channel pulse response (response has length N) x_k is the transmitted (PAM-4) symbol at time k n_k is the noise value at time k	$f_k = \sum_{i=0}^{N_w-1} w_i\, y_{k-i}$ w_i is a coefficient of the feed-forward equalizer (N_w taps)	$g_k = x_{k-d} + \sum_{j=1}^{N_b} b_j\, x_{k-d-j}$ d is $d_h + d_w$ where d_w is the FFE delay b_j is a coefficient of the feedback equalizer (N_b taps)

System model, matrix form



Note: Vectors and matrices are denoted by **bold face** type

Channel output	FFE output	Target signal
$\boldsymbol{y}_k = \boldsymbol{x}_k\boldsymbol{H} + \boldsymbol{n}_k$ $\boldsymbol{x}_k$ is a row vector of the $N + N_w - 1$ most recently transmitted symbols x_k to x_{k-N-N_w+2} $\boldsymbol{h}$ is the vector $[h_{-d_h}, ..., h_0, ..., h_{N-d_h-1}]$ $\boldsymbol{H}$ is a diagonal-constant (Toeplitz) matrix — first column is $\boldsymbol{h}$ followed by $N_w - 1$ zeros — first row is h_{-d_h} followed by $N_w - 1$ zeros $\boldsymbol{n}_k$ is a row vector of the N_w most recent noise values n_k to n_{k-N_w+1}	$f_k = \boldsymbol{y}_k\boldsymbol{w}$ $\boldsymbol{y}_k$ is a row vector of the N_w most recent FFE inputs y_k to y_{k-N_w+1} $\boldsymbol{w}$ is a column vector of the FFE coefficients w_0 to w_{N_w-1}	$g_k = \boldsymbol{x}_{k-d}\boldsymbol{p}$ $\boldsymbol{x}_{k-d}$ is a row vector of delayed symbols x_{k-d} to x_{k-d-N_b} $\boldsymbol{p}^T$ is $[1 \;\; \boldsymbol{b}^T]$ where a T exponent denotes the matrix transpose $\boldsymbol{b}$ is a column vector of feedback coefficients b_1 to b_{N_b}
		

Definition of mean-squared error

$E\left[e_k^2\right] = E[(f_k - g_k)^2] = E\left[f_k^2\right] + E\left[g_k^2\right] - 2E[f_k g_k]$ where $E[x]$ is the expected value of random variable x

$E\left[e_k^2\right] = \boldsymbol{w}^T\boldsymbol{R}_{yy}\boldsymbol{w} + \boldsymbol{p}^T\boldsymbol{R}_{xx}\boldsymbol{p} - 2\boldsymbol{w}^T\boldsymbol{R}_{yx}\boldsymbol{p}$

where $R_{yy} = E\left[\boldsymbol{y}_k^T\boldsymbol{y}_k\right] = E\left[\left(\boldsymbol{H}^T\boldsymbol{x}_k^T + \boldsymbol{n}_k^T\right)(\boldsymbol{x}_k\boldsymbol{H} + \boldsymbol{n}_k)\right] = \sigma_X^2\boldsymbol{H}^T\boldsymbol{H} + \boldsymbol{R}_{nn}$ where $\boldsymbol{R}_{nn} = E\left[\boldsymbol{n}_k^T\boldsymbol{n}_k\right]$

$R_{xx} = E\left[\boldsymbol{x}_{k-d}^T\boldsymbol{x}_{k-d}\right] = \sigma_X^2\boldsymbol{I}_p$ where $\boldsymbol{I}_p$ is the N_p-by-N_p identity matrix

$R_{yx} = E\left[\boldsymbol{y}_k^T\boldsymbol{x}_{k-d}\right] = E\left[\left(\boldsymbol{H}^T\boldsymbol{x}_k^T + \boldsymbol{n}_k^T\right)\boldsymbol{x}_{k-d}\right] = \sigma_X^2\boldsymbol{H}_p^T$ where $\boldsymbol{H}_p$ is rows $d+1$ to $d+N_b+1$ from $\boldsymbol{H}$

Let $\boldsymbol{R} = \boldsymbol{R}_{yy}/\sigma_X^2 = \boldsymbol{H}^T\boldsymbol{H} + \boldsymbol{R}_{nn}/\sigma_X^2$

$$\boxed{E\left[e_k^2\right] = \sigma_X^2\left(\boldsymbol{w}^T\boldsymbol{R}\boldsymbol{w} + \boldsymbol{p}^T\boldsymbol{p} - 2\boldsymbol{w}^T\boldsymbol{H}_p^T\boldsymbol{p}\right)}$$

Recall that $\boldsymbol{p}^T$ = [1 $\boldsymbol{b}^T$]. Let where $\boldsymbol{h}_0$ be row $d+1$ from $\boldsymbol{H}$ and $\boldsymbol{H}_b$ be rows $d+2$ to $d+N_b+1$ from $\boldsymbol{H}$

$$\boxed{E\left[e_k^2\right] = \sigma_X^2\left(\boldsymbol{w}^T\boldsymbol{R}\boldsymbol{w} + 1 + \boldsymbol{b}^T\boldsymbol{b} - 2\boldsymbol{w}^T\boldsymbol{h}_0^T - 2\boldsymbol{w}^T\boldsymbol{H}_b^T\boldsymbol{b}\right)}$$

Minimum mean-squared error (MMSE) optimization

Find $\boldsymbol{w}$ and $\boldsymbol{b}$ that minimize mean-squared error subject to an equality constraint on $\boldsymbol{w}^T\boldsymbol{h}_0^T$ (amplitude of the equalized pulse).

Use the method of Lagrange multipliers. Begin with the Lagrange function.

$\mathcal{L}(\boldsymbol{w},\boldsymbol{b},\lambda) = u(\boldsymbol{w},\boldsymbol{b}) + \lambda v(\boldsymbol{w})$ where λ is the Lagrange multiplier.

where $u(\boldsymbol{w},\boldsymbol{b}) = E\left[e_k^2\right]$ Minimize the mean-squared error ...

$v(\boldsymbol{w}) = -2\sigma_X^2(\boldsymbol{w}^T\boldsymbol{h}_0^T - 1)$ ... subject to $v(\boldsymbol{w}) = 0$ ($\boldsymbol{w}^T\boldsymbol{h}_0^T = 1$)

$$\mathcal{L}(\boldsymbol{w},\boldsymbol{b},\lambda) = \sigma_X^2\left(\boldsymbol{w}^T\boldsymbol{R}\boldsymbol{w} + 1 + \boldsymbol{b}^T\boldsymbol{b} - 2\boldsymbol{w}^T\boldsymbol{h}_0^T - 2\boldsymbol{w}^T\boldsymbol{H}_b^T\boldsymbol{b} - 2\lambda\boldsymbol{w}^T\boldsymbol{h}_0^T + 2\lambda\right)$$

Take the partial derivatives of the Lagrange function with respect to $\boldsymbol{w}$, $\boldsymbol{b}$, and λ and set them to 0.

$$\left.\begin{aligned}
\frac{\mathcal{L}(\boldsymbol{w},\boldsymbol{b},\lambda)}{d\boldsymbol{w}} &= 2\sigma_X^2\left(\boldsymbol{R}\boldsymbol{w} - \boldsymbol{h}_0^T - \boldsymbol{H}_b^T\boldsymbol{b} - \lambda\boldsymbol{h}_0^T\right) = 0 \\
\frac{\mathcal{L}(\boldsymbol{w},\boldsymbol{b},\lambda)}{d\boldsymbol{b}} &= 2\sigma_X^2(\boldsymbol{b} - \boldsymbol{H}_b\boldsymbol{w}) = 0 \\
\frac{\mathcal{L}(\boldsymbol{w},\boldsymbol{b},\lambda)}{d\lambda} &= 2\sigma_X^2(-\boldsymbol{h}_0\boldsymbol{w} + 1) = 0
\end{aligned}\right\}$$

This is a system of $N_w + N_b + 1$ equations with $N_w + N_b + 1$ unknowns.

$$\begin{bmatrix} \boldsymbol{R} & -\boldsymbol{H}_b^T & -\boldsymbol{h}_0^T \\ -\boldsymbol{H}_b & \boldsymbol{I}_b & \boldsymbol{z}_b^T \\ \boldsymbol{h}_0 & \boldsymbol{z}_b & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{w} \\ \boldsymbol{b} \\ \lambda \end{bmatrix} = \begin{bmatrix} \boldsymbol{h}_0^T \\ \boldsymbol{z}_b^T \\ 1 \end{bmatrix}$$

where $\boldsymbol{I}_b$ is the N_b x N_b identity matrix and $\boldsymbol{z}_b$ is a row vector of N_b zeros

MMSE optimization, continued

Solve the system of equations. This is readily done by computer and the derivation is not continued beyond this point.

$$\begin{bmatrix} \boldsymbol{w} \\ \boldsymbol{b} \\ \lambda \end{bmatrix} = \begin{bmatrix} \boldsymbol{R} & -\boldsymbol{H}_b^T & -\boldsymbol{h}_0^T \\ -\boldsymbol{H}_b & \boldsymbol{I}_b & \boldsymbol{z}_b^T \\ \boldsymbol{h}_0 & \boldsymbol{z}_b & 0 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{h}_0^T \\ \boldsymbol{z}_b^T \\ 1 \end{bmatrix}$$

It can shown that the solution for the Lagrange multiplier λ is the mean-squared error normalized to σ_X^2. However, this result is not used since it may not be the correct value when limits are later imposed on $\boldsymbol{b}$ and $\boldsymbol{w}$.

MMSE optimization with fixed feedback coefficients

Given $\boldsymbol{b}$, find $\boldsymbol{w}$ that minimizes mean-squared error subject to the constraint $\boldsymbol{w}^T\boldsymbol{h}_0^T = 1$.

Recall that $\boldsymbol{p}^T$ = [1 $\boldsymbol{b}^T$]. Use the method of Lagrange multipliers as before.

$$\mathcal{L}(\boldsymbol{w},\boldsymbol{b},\lambda) = u(\boldsymbol{w},\boldsymbol{b}) + \lambda v(\boldsymbol{w})$$

where $u(\boldsymbol{w},\boldsymbol{b}) = E\left[e_k^2\right]$

$$v(\boldsymbol{w}) = -2\sigma_X^2(\boldsymbol{w}^T\boldsymbol{h}_0^T - 1)$$

$$\mathcal{L}(\boldsymbol{w},\boldsymbol{b},\lambda) = \sigma_X^2\left(\boldsymbol{w}^T\boldsymbol{R}\boldsymbol{w} + \boldsymbol{p}^T\boldsymbol{p} - 2\boldsymbol{w}^T\boldsymbol{H}_p^T\boldsymbol{p} - 2\lambda\boldsymbol{w}^T\boldsymbol{h}_0^T + 2\lambda\right)$$

Take the partial derivatives of the Lagrange function with respect to $\boldsymbol{w}$ and λ and set them to 0.

$$\left.\begin{aligned} \frac{\mathcal{L}(\boldsymbol{w},\lambda)}{d\boldsymbol{w}} &= 2\sigma_X^2\left(\boldsymbol{R}\boldsymbol{w} - \boldsymbol{H}_p^T\boldsymbol{p} - \lambda\boldsymbol{h}_0^T\right) = 0 \\ \frac{\mathcal{L}(\boldsymbol{w},\lambda)}{d\lambda} &= 2\sigma_X^2(-\boldsymbol{h}_0\boldsymbol{w} + 1) = 0 \end{aligned}\right\}$$

Solve the system of $N_w + 1$ equations with $N_w + 1$ unknowns.

$$\begin{bmatrix} \boldsymbol{R} & -\boldsymbol{h}_0^T \\ \boldsymbol{h}_0 & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{w} \\ \lambda \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_p^T\boldsymbol{p} \\ 1 \end{bmatrix} \longrightarrow \boxed{\begin{bmatrix} \boldsymbol{w} \\ \lambda \end{bmatrix} = \begin{bmatrix} \boldsymbol{R} & -\boldsymbol{h}_0^T \\ \boldsymbol{h}_0 & 0 \end{bmatrix}^{-1}\begin{bmatrix} \boldsymbol{H}_p^T\boldsymbol{p} \\ 1 \end{bmatrix}}$$