# Channel Characterization for Evaluating Advanced Modulation Systems

Will Bliss
Office of the CTO
Broadcom Corp.

IEEE 802.3bs Task Force Workshop

San Diego, CA

July 15, 2014

# OUTLINE

- Assumptions, Motivations and Goals
- A theoretical and practical look at PAM vs. DMT
- History and speculation of PAM vs. DMT for fiber-optics
- What is Signal, Noise, and Distortion?
- Proposed method to characterize channels and share data
- Future work

# ASSUMPTIONS WITH MOTIVATIONS

- **Low (lowest?) cost and power electro-optical parts**
  - So only Direct Detection (DD) of optical Intensity
  - I.e., since not considering long haul, then not considering ‘coherent,’ where the optical field phase (aka carrier) can be controlled (modulated) at the TX, and ‘tracked’ (detected) at the RX
- **Low (lowest?) power electrical solutions**
  - Electronics power is first order proportional to ‘Baud Rate,’ so we can’t afford the power of Transmitting and Receiving frequencies that deliver little or no information
    - So we limit ourselves to Band-Width limited systems where SNR(f) >> 0dB for all frequencies used
    - E.g., no ‘integer oversampling’ (a.k.a. fractionally spaced equalizers, etc.)
  - But will consider DSP to mitigate certain electro-optics issues
    - Limited trade of electrical power vs. optical costs

# GOALS

- **Create and share 'channel data' between different advocacy groups**
  - Enough to allow the science of reproducing other's experiments
- **Create and share 'channel models' between different advocacy groups**
  - Enough to allow 'optimization' of different systems for the fairest comparisons
- **Without too much extra work (data collection, experiments, etc.)**
  - A difficult challenge, given that PAM-M and DMT have different TX power spectrums and different TX probability distributions

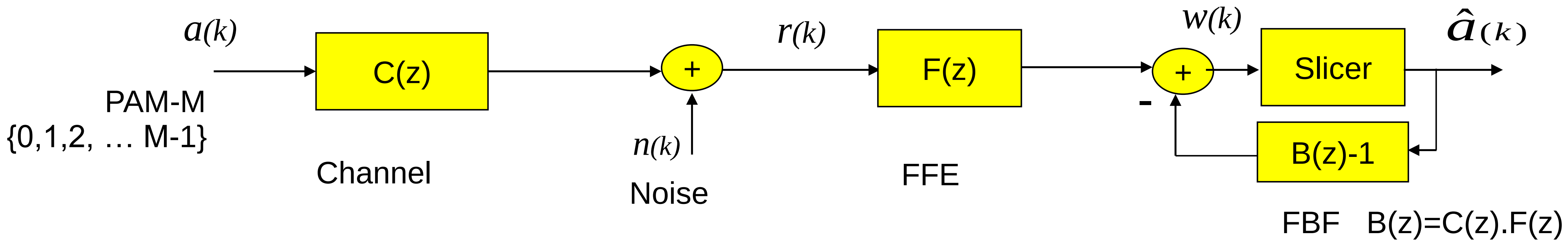
# PAM VS. DMT?

- **The communication theory literature deals with this choice as ‘Single Carrier’ vs. ‘Multi-Carrier’**
  - Pulse Amplitude Modulation (PAM) and Quadrature Amplitude Modulation (QAM) are the most common examples of ‘Single Carrier’
  - Discrete Multi Tone (DMT) and Orthogonal Frequency Division Multiplexing (OFDM) are the most common example of ‘Multi-Carrier’
- **Both are very general categories that leave a host of details open**
  - Kind of like saying, “We will send data with Photons.” This still leaves many important variants to specify!
- **And there are a near infinite number of such variants**
  - Technically, PAM-M is a trivial subset of DMT using N=1 frequency bins
  - So fully optimized DMT can NOT be inferior to PAM-M (else it would optimize to be PAM-M)

# PULSE AMPLITUDE MODULATION (PAM-M)

- **A type of ‘Base-Band’ modulation (== Single Carrier at DC)**
- **Without loss of generality, performance fundamentally identical to QAM**
  - Difference is band-pass vs. baseband channels, ignored herewith
- **Generally with uniformly spaced amplitude levels**
  - Without loss of generality, think of PAM-M as the set {0,1,2, …, M-1}
  - And NRZ == PAM-2, on {0,1}
  - Typical for base-band implementations, DC and very low frequencies are normally finessed with various well known techniques
  - PAM generalizes to non-uniform levels, which have been suggested for very high RIN dominated channels [bhoja_01_0112_NG100GOPTX]
- **Transmitters can be ultra-simple**
  - Only the small number of M levels, so no high resolution multi-bit DACs are required

# PAM WITH DECISION FEEDBACK EQUALIZER (DFE)

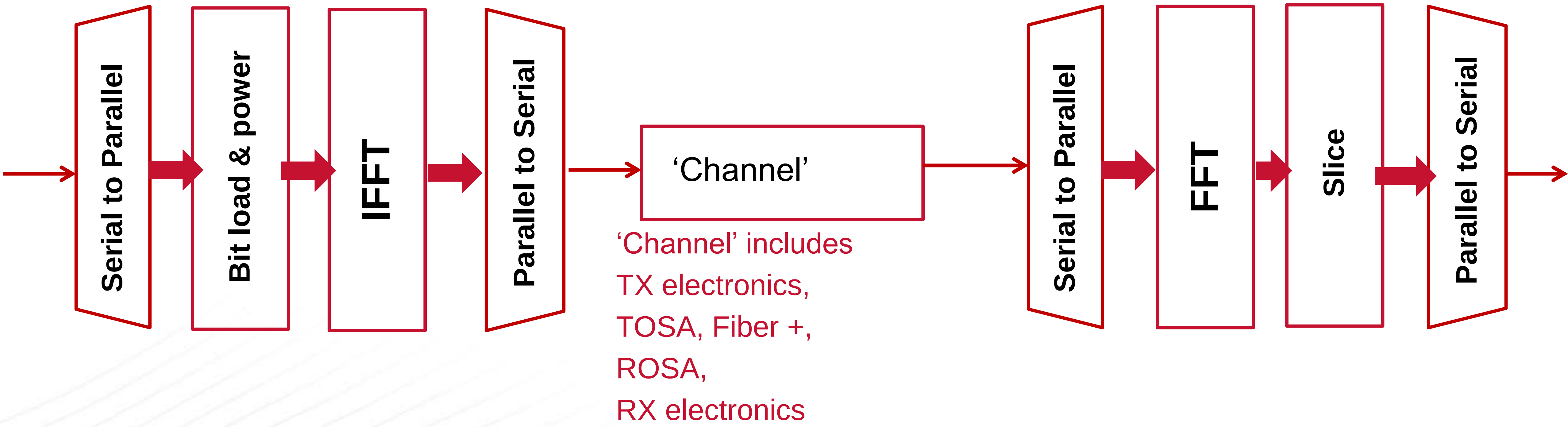



- **Optimal DFE both 'whitens' the noise at *w(k)* and transforms the signal to minimum phase**
- **The performance of an MMSE optimized DFE achieves the Salz *SNR_dB = mean{10*log10[1+SNR(f)]}* over the Nyquist Band**
- **The ZFE-DFE solution achieves the geometric mean of the SNR(f), which for high SNR converges to the Salz SNR**
- **In many SERDES applications, the Feed Forward Equalizer (FFE) function is performed solely by, or in conjunction with, an analog Continuous Time Filter (CTF)**
- **'Very Good' channels' allow ultra-simple inverting equalizer, F(z) ~= 1/C(z), so no FBF (no DFE) required**

# DISCRETE MULTI TONE (DMT)

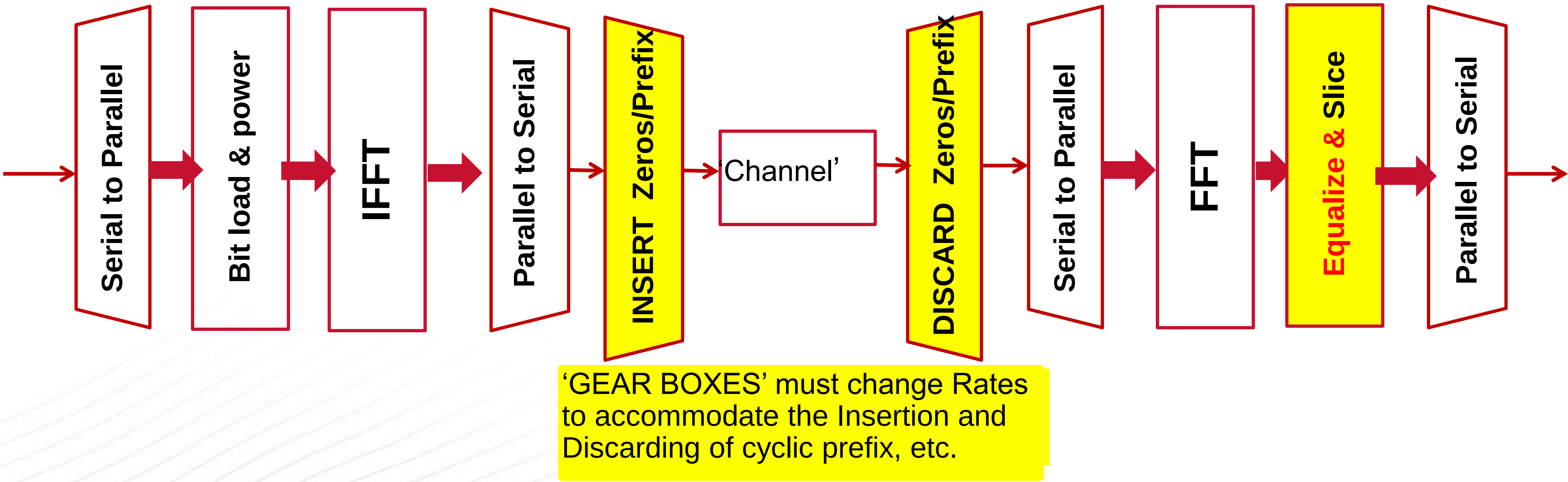
- **DMT performs IFFTs on blocks of data to be transmitted, and performs FFTs on blocks of received data**
- **Because the ‘blocks of data’ are essentially rectangular windows in time, the DMT isn’t exactly ‘multi-carrier’**
  - Frequency ‘bins’ are not pure tones, but are SINC(f) shape in frequency, so they overlap significantly in frequency
  - But the ‘bins’ are orthogonal when synchronized to the block, so they don’t interfere with each other in a synchronized receiver
  - The connection to Fourier (and FCC) ‘frequency’ is maintained well enough for most uses
- **DMT maps simple ‘integer / digital user data’ into real (continuous, like analog) values to be transmitted**
  - So high resolution (multi-bit, approaching ‘continuous’) DAC and TX are required

# DISCRETE MULTI -TONE (DMT-N)



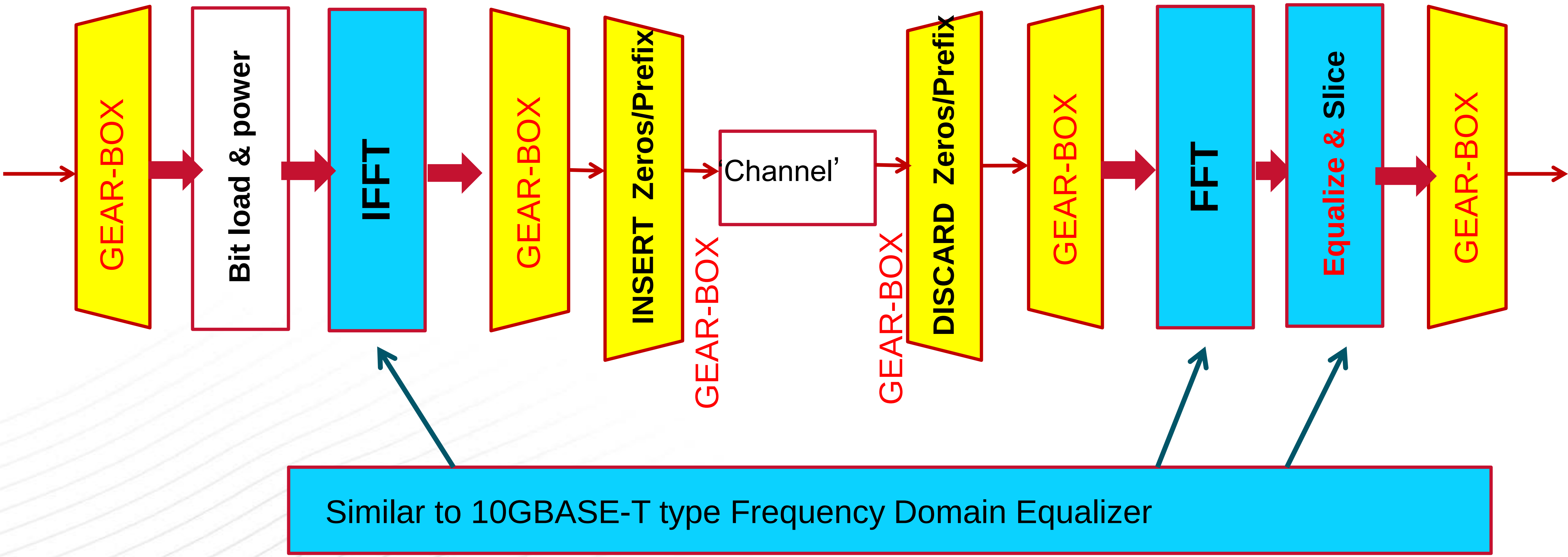

- **If the complete 'channel' response is the 'unit pulse' (no ISI), then the IFFT * FFT block operation is the identity matrix, so 'lossless' (ignoring any noises)**
- **Without loss of generality we can consider each 'frequency bin' (each bit loading) as either QAM or PAM. System performance is nearly identical and either serves our illustrations**
  - We follow the literature with the order IFFT ➔ FFT, so think of frequency domain data bins (input and output) and time domain values in the channel

# DMT WITH CYCLIC PREFIX, ETC FOR ISI CHANNEL




- **Typical applications have 'some ISI' of length L in the channel**
- **Typical 'lowest complexity' fix is to prepend L Bauds of 'cyclic prefix' (non-information carrying) to each block of N Baud samples**
  - Which converts the FFT based 'cyclic convolution' into an effective linear convolution
  - Each of N frequency bins are 'equalized' with complex multipliers

# DMT, DESCRIBED IN 802.3 TERMINOLOGY

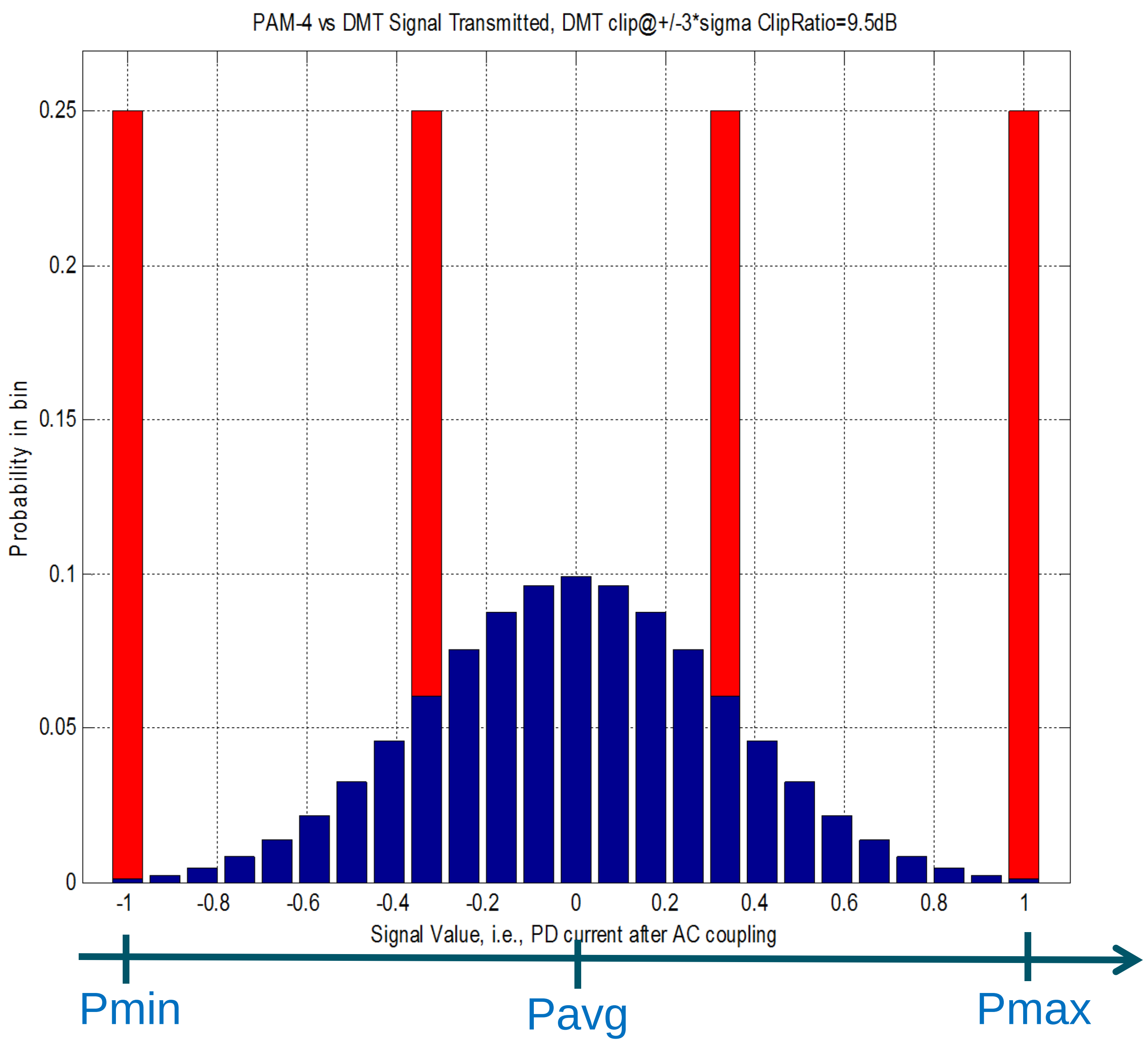



- **The DMT achieves channel equalization with a Frequency Domain approach, very similar to that used in 10GBASE-T implementations**
- **Except the DMT proposals runs at over 70 times higher speed (throughput)**

# DMT DESIGN TRADEOFFS

- **Typically the ‘DC bin’ and the ‘Nyquist bin’ are zeroed (no information) because of practical difficulties**
- **Typically one or more bins are dedicated to ‘pilot tones’ to accomplish Clock Data Recovery (CDR)**
- **Typically many non-information Bauds are inserted per block of N Bauds for ‘cyclic prefix’ or similar**
- **All three of these above push towards large ‘N’, the number of Baud samples in the FFT engines, but …**
- **Large N increase the computational burden per information bit**
- **Large N increases the number of ‘constants’ which must be learned and saved. E.g., expect ~1,000 Bytes of storage**
- **Large N increases the Peak to Average Ratio (PAR) of the transmitted and received signals**
  - Typical implementations use designed clipping at the TX to avoid further loss of average signal variance (a.k.a. power in communications)

# DMT TX 'SIGNAL VARIANCE' AND CLIPPING




- **Red = PAM-4 probability**
- **Blue = DMT example with moderate clipping at +/- 3*sigma**
  - 'Clipping ratio' = 9.5dB
- **Mean time to 'clipping' is about 370 Bauds, so average more than one clip per Block of N=512 Baud samples.**
  - Many blocks will have multiple clippings
- **The 'Signal Variance' (which is communication theory TX power) is 7 dB lower than that of PAM-4**
- **Note that the laser has the same peak-peak power range and equal average power**

# THEORY OF COMMUNICATION ; CAPACITY

BROADCOM

- **Consider Linear Time Invariant (LTI) channels with Additive White Gaussian Noise (AWGN)**
  - This is the main theoretical development in the literature
  - For additive colored Gaussian noise, it's easy to 'whiten' and apply the theory
- **The Shannon-Hartley Capacity of such channels (for asymptotic zero probability of error) for Band Width limited channels is**

$$C_{SH} = \int_0^{BW} log_2 \left(1 + \frac{S(f)}{N(f)}\right) df$$

- **Surprisingly, this Capacity is just a scaled version of the DFE's Salz SNR,** ***$C_{SH}$ = (Fs/6.02)(bit/dB) * Salz_SNR_dB***
- **Conclusion is that IF you have high SNR, then you are motivated to try and send more than one bit/Baud**

# MAXIMIZING CAPACITY

- **First proof of maximizing capacity was with the 'water filling' algorithm, which apportions a given finite *average* TX power**
  - Whose 'frequency selectivity' is akin to multi-carrier
  - Which for some time was thought to be a requirement to approach capacity
- **But further development of PAM-DFE showed 'surprising' results**
  - [Cioffi, Dudevoir,Eyuboglu, Forney, IEEE Tr. Communication Thy., Oct. 95] showed that an unbiased MMSE-DFE is a canonical (lossless) receiver
    "Finally, at the optimized symbol rate, there is no distinguishable performance difference between a flat transmit spectrum and an optimized spectrum on a wide range of channels with ISI."
    "There are also differences in implementation and system issues between single-tone and multitone systems, such as how *filtering* is implemented, how the system is *adapted*, *delay*, sensitivity to other types of *distortion*, and so forth. … Since, as we have seen, there will be essentially no difference in maximum achievable SNR performance between these two classes of systems, particularly when used with powerful codes, the choice between them will come down to other factors, such as these."
    "Simulation results suggest an even stronger result: on typical ISI channels, a non-optimized flat transmit spectrum yields near-optimal MMSE-DFE performance down to rather low SNR, …"

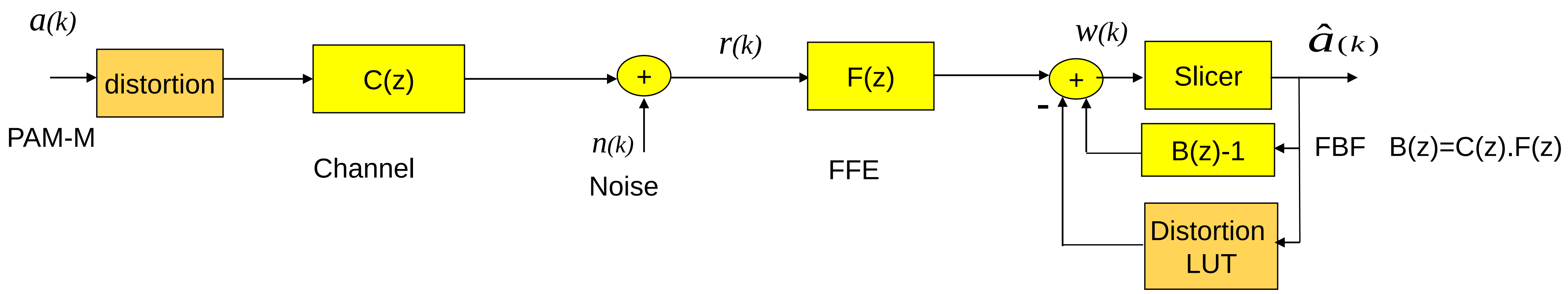
# CONCLUSIONS ON CAPACITY FOR DMT AND PAM

- **DMT offers no theoretical ‘spectral efficiency’ advantages over PAM-DFE on any reasonable channels we’ll consider**
  - We don’t (shouldn’t) have a large number of spectral / SNR(f) nulls
  - We don’t want to operate over significant bands where SNR(f) is low
  - A flat TX power spectrum is sufficient for optimal single carrier
- **Our peak constrained transmitters give a 6-7dB disadvantage to DMT vs. PAM**
  - For reasonably ‘good’ (flat SNR(f)) channels, DMT can’t overcome this disadvantage
- **For high ISI channels, or more accurately, channels with high variation in SNR(f),**
  - The DMT system must ‘water fill’ allocate TX power to try to gain back some of the loss from lower TX power
  - The PAM-M system must go beyond ‘inverting equalizer,’ to architectures that whiten the noise (such as DFE, THP, MLSD, etc.)
  - We find the implementation power / cost of DMT is ~2x higher than PAM, so it makes sense to search beyond ‘Vanilla PAM with inverting equalizer,’ IF we’re going to support these ‘poor channels’

# HISTORY OF PAM VS. DMT, WITH SPECULATIONS

- **PAM advocates have generally been assuming 'quite good' channels**
  - Relatively high BW
  - Often using MZM, with little or no distortion
  - With fairly good RIN, such that there is little spread of SNR across different 'levels'
  - All the channels supported with simple 'inverting equalizer' + slicer for the RX (no DFE or other advanced detector, etc.)
- **DMT advocates have been working on 'quite difficult' channels**
  - Quite low BW. E.g., using 10Gbps DML components to achieve 100Gbps
  - Net SNR(f) graphs show surprising difficulties
    - Some show a very large number of 'notches' with a comb like structure. We speculate that these are due to time-interleaving errors in the DAC and/or the ADC
    - Some (all) show SNR(f) going to 0dB. We speculate these are again due to DAC and/or ADC , or are from a RIN peaking effect, or are simply from severe ISI, or?
  - Possibly from significant DML distortions
    - DMT handles certain impulse-like distortions well (like clipping)
    - Note that some DML PAM systems showed 'asymmetric eyes' that are explained by distortions
- **We should do some science and work on the same channels!**
  - Share channel data and channel models
  - Reproduce work and optimize systems

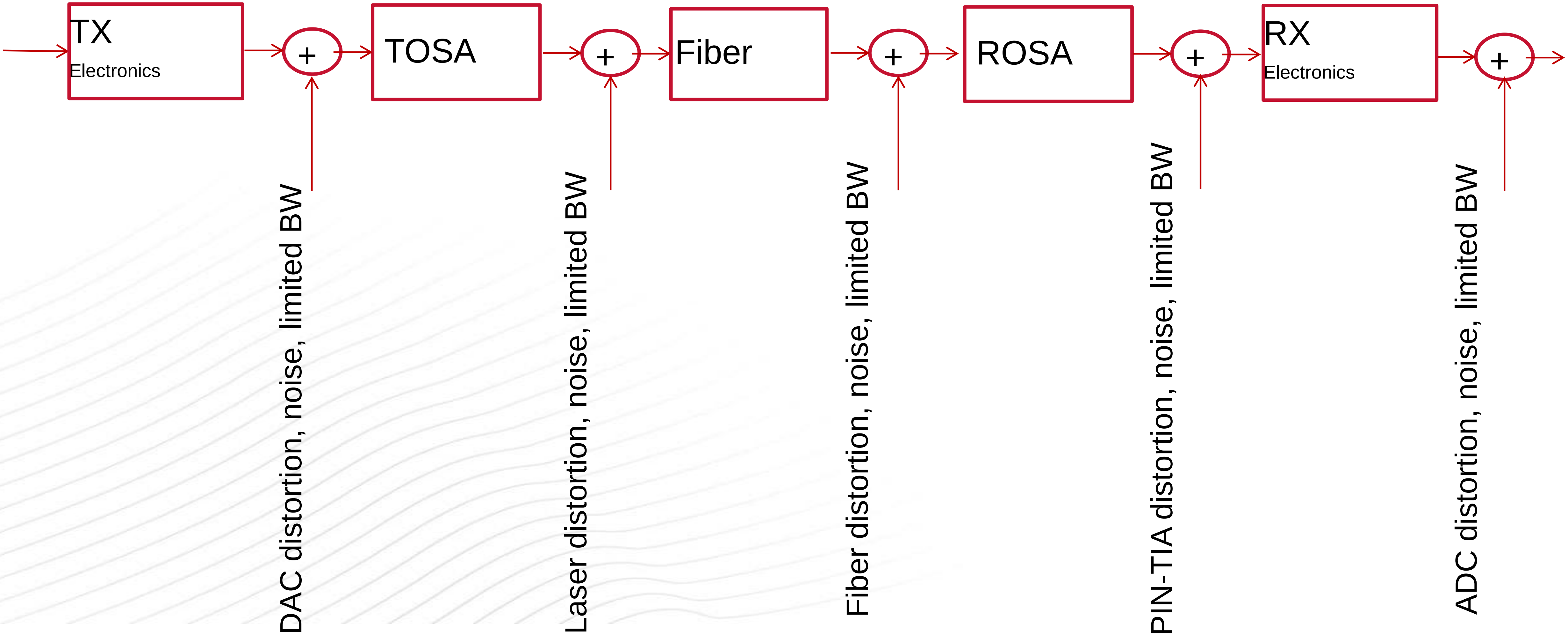
# PAM WITH NON-LINEAR DECISION FEEDBACK EQUALIZER (DFE)




- **The best approach to handling distortion is to model it and incorporate it into the system as a known a priori**
  - Not to 'smear it' over a block of N symbols as in DMT
- **A very simple and effective such architecture is the non-linear DFE, which cancels the trailing distortion terms and optimally sets the decision thresholds**
  - See [Winters, Kasturia, Jounal of Lightwave Tech., July 1992]
  - So the system performance goes with the {Signal / Noise}, effectively ignoring the distortion energy

# THE 'CHANNEL' IS MORE COMPLEX THAN JUST 'RX SENSITIVITY'

- **Distortions and some noises are 'Data Dependent' (not additive)**
- **Seems too difficult to completely specify each component?**



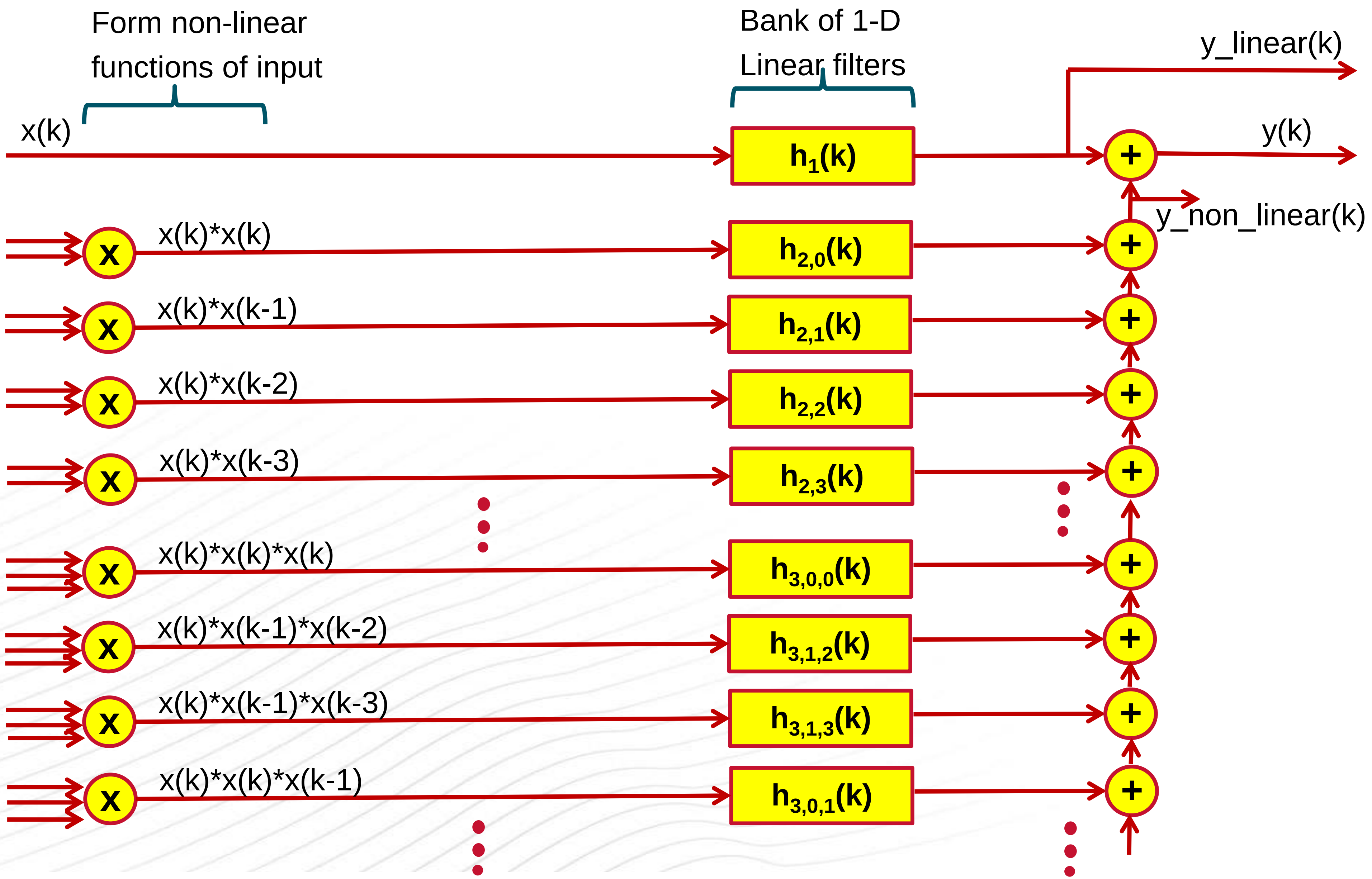
# WHAT IS SIGNAL, NOISE, AND DISTORTION?

- **Everyone thinks they know, but different applications and individuals use the terms differently, so DEFINE here:**
- **Noise = the non-repeatable portion of observed waveforms (experiments) when we believe all the known experimental conditions are identical (e.g., the same data pattern)**
  - Probably need to include temperature, pressure, humidity, ‘voltage’, etc., as ‘controls
  - Define {Signal + Distortion} as the result of averaging out the Noise
  - Note that the Noise can also be broken into ‘stationary’ and ‘time varying’
- **Signal == the portion of the {Signal + Distortion} above that is ‘fit’ by a Linear Time Invariant (LTI) model**
  - Say of length L Baud samples
  - The effect of the LTI systems is Inter Symbol Interference (ISI)
  - Sometimes called the ‘Linear part of the Signal’
- **Distortion = the portion of the {Signal + Distortion} that is NOT fit by the LTI model above**
  - Sometimes called the ‘Non- Linear part of the Signal’

# PROPOSAL FOR CHANNEL MODELING

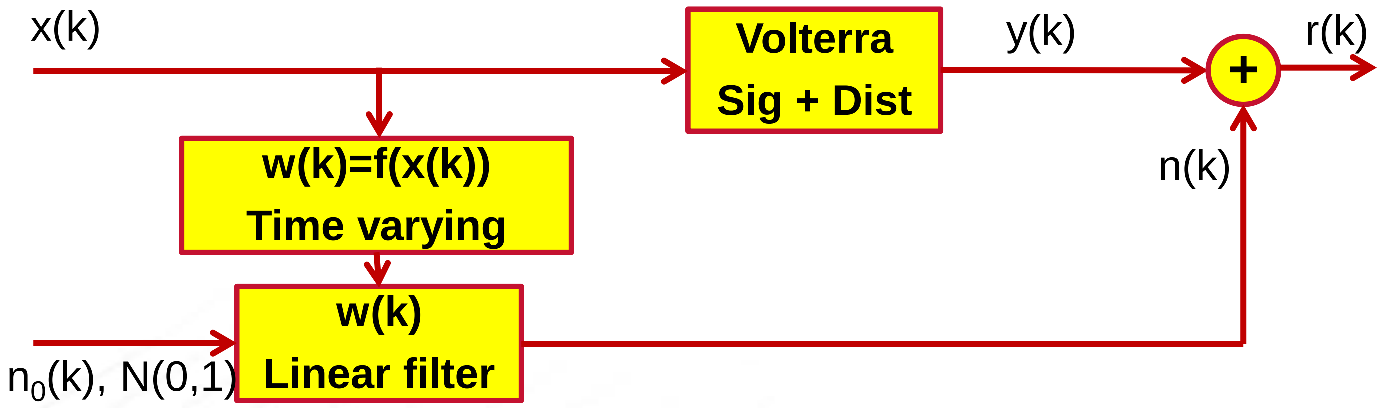
- **Choose a rich but relatively short input pattern, x(k), to excite the system**
  - With support from physics modeling for expected length of correlations and dependencies
  - Repeat the pattern P times (as possible, to support easy separation of the noise)
- **Collect observations (lab measurements or optical simulations) and take the mean over the repeated periods to form estimates of both the {noise} and the {signal + distortion}**
- **Construct a Volterra filter to synthesize the {signal + distortion} from the known input data**
  - The Volterra filter is almost always impractical for product implementation, but here it is only used for studying (modeling) channels
  - The Volterra filter can be stimulated with new (synthetic) excitation data without going back to more data collection
- **Construct a time-varying (input data dependent) filter to synthesize the data dependent and colored noise**

# VOLTERRA SYNTHESIS MODEL OF SIGNAL AND DISTORTION




- **Volterra filter used to synthesize the observed signal and distortion**
- **x(k) is the known data sent to 'TX'**
- **y(k) is fit to the observed data**
- **Here viewed as a bank of linear 1-Dimensional LTI filters, one for each 'non-linear input'**
- **Allows solving with standard MMSE linear algebra**

# SYNTHESIS MODEL WITH TIME VARYING NOISE




- **The synthesis model is most useful when augmented with the ability to synthesize noise that matches the time varying (data dependent) statistics of the observed data**
- **This can be achieved by filtering a random noise source with a time-varying linear filter**
- **The observed noise distributions are very close to Gaussian, especially at the high raw BERs supported by high coding gain FEC**

# FUTURE WORK

- **Develop standardized methods to describe (and share) channel models that include the distortion(s) and time-varying noise(s)**
  - Can we share an input excitation, x(k), or do we need two separate?
  - PAM-4 performance can be well described by measurements with ~PAM-64, which we expect will well predict DMT performance
- **Start working on the same channels!**
  - We need models of both ‘poor’ and ‘good’ channels to experiment with
- **PAM advocates to demonstrate advanced RX performance and cost on a ‘poor channel’ modeled above (e.g., with ISI and distortion)**
- **DMT advocates to demonstrate performance and cost on a ‘poor channel.’ And cost on a ‘good channel’?**
- **Propose an electrical Reference TX and a Reference RX to define the Channel (pass/fail)**
- **Answer the question “How poor of channels does it make sense to support?”**
  - The goal is a cost and power effective product, not just a technical possibility
  - See 10GBASE-T history

# Thank you