# Transmit Pre-emphasis for High-Speed Time-Division-Multiplexed Serial-Link Transceiver

Vladimir Stojanovic, George Ginis, Mark A. Horowitz
Department of Electrical Engineering, Stanford University, CA 94305, USA

***Abstract*-Time Division Multiplexing (TDM) must be employed in multi-GSa/s transceivers in order to overcome on-chip clock frequency limitations. This paper describes a transmit pre-emphasis filter for a multi-level transceiver making use of TDM. The possible applications of such a transceiver include serial links and chip-to-chip communication. The requirement of very low probability of error in the absence of coding, and the need for an adaptive solution impose a peak transmit power constraint. The TDM system is mapped to a Multiple-Input-Multiple-Output (MIMO) system, and the noise sources are analyzed. The design of the pre-emphasis filter is shown to be a non-convex optimization problem, whose optimal solution is very difficult to obtain. Still, sub-optimal solutions are derived in closed form and adaptive implementations are described. Simulation results using parameters obtained from an experimental testbed indicate that these sub-optimal solutions actually achieve very good performance.**

## I. Introduction

Increasing demand for high-bandwidth chip-to-chip communication is driving the development of systems that communicate at multi-Gb/s rates over cables or backplane traces. In addition to challenges in IC design to sustain these rates, the communication channel is entering the region where Inter-Symbol Interference (ISI) becomes the dominant limiting factor in the overall performance of the system. The major sources of ISI in the channel are parasitic RLC filtering at both the transmitter and receiver sides, skin effect and dielectric loss in the cable or backplane traces and reflections from inductive and capacitive discontinuities from the chip package, connectors and via stubs. In order to achieve very high bit rates, multi-level transmission and equalization must be used.

Data transmission systems that rely on the assumption of a stationary channel have been recently demonstrated [1], [2] at multi-Gb/s rates. These systems employed 4-PAM modulation, used 2-bit Digital-to-Analog Converters (DACs) and Analog-to-Digital Converters (ADCs), and included pre-programmed analog filters for channel equalization. This paper describes equalization techniques with adaptive implementations that have been developed for and tested on an 8 GSa/s serial-link transceiver employing either 2-PAM or 4-PAM [3].

As shown in Fig. 1, the transmitter consists of eight time-multiplexed 8-bit DACs, in groups of two, clocked from the transmitter PLL through the phase adjusters. The receiver consists of eight time-multiplexed 4-bit ADCs. The phase adjusters in the receiver synchronize each ADC (with a sampling rate of 1GSa/s) with one of the transmitting DACs, thus achieving an aggregate sampling rate of 8GSa/s.

Research supported by the MARCO Interconnect Center and gifts from Alcatel, Fujitsu, Samsung, France-Telecom, IBM, Voyan, Sony and Telcordia.

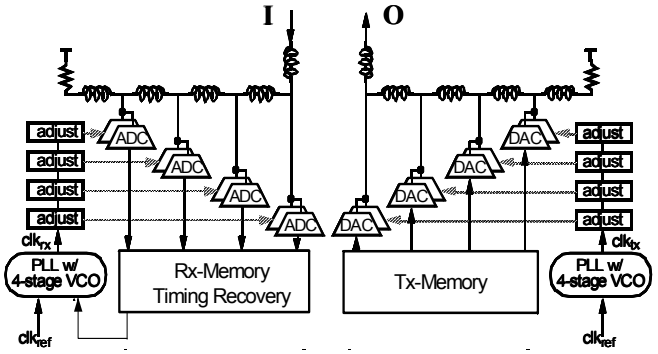


Fig. 1. Transceiver block diagram.

The fundamental reason for performing TDM is to avoid the on-chip clock frequency limit in current CMOS technology. However, this significantly increases the parasitic RC filtering at the transmitter output/receiver input, thus reducing the useful bandwidth of the link. An attempt is made to "distribute" the parasitic capacitance by insertion of inductors between each pair of transmitters and receivers in order to form a lumped LC transmission line in a manner analogous to distributed amplification [4]. This form of analog equalization extends the useful bandwidth of the link, effectively without any noise enhancement penalty.

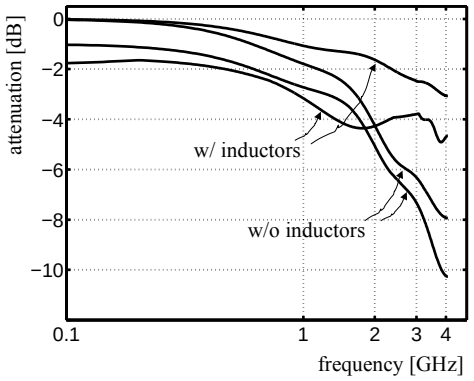


Fig. 2. Measured transmitter frequency response (envelopes of all 8 DACs) with and without inductors.

A comparison of the lower and upper bounds of the "distributed" vs. the "non-distributed" transmitter frequency response is given in Fig. 2. The data represents the FFT of the pulse response captured by a sampling oscilloscope after one meter of coaxial cable from point **O**, Fig. 1, in the experimental setup presented in [3]. Large variations in the bondwire inductors result in significant differences between the TDM sub-channels. Instead of exhibiting the behavior of a lumped LC transmission line, the frequency response is

dominated by second-order peaking. Evidently, the variation between transmitter responses is much smaller in the case without inductors than in the case with inductors. However, the useful bandwidth increase in the "distributed" case is apparent.

The "extra resolution" of DACs and ADCs enables equalization and correction of "static" noise. The resolution of transmitter DACs and receiver ADCs is limited by the current chip technology, and is *typically much higher for the DAC than for the ADC*. This dictates the use of transmit pre-emphasis filtering as opposed to the traditionally used receiver equalization. Due to the very high data rates, coding techniques cannot be applied, but still the system has to be designed for a probability of error smaller than $10^{-20}$. These characteristics of the system make equalization a critical issue.

The general system description is given in Section II. Section III first describes the equalizer design problem, where a peak power constraint applies instead of the typical average power constraint. Optimal solutions are shown to have prohibitive computational cost, so closed form sub-optimal solutions are given and adaptive implementations are described. Section IV presents the results of simulations, where the pulse response data are obtained from an experimental testbed.

## II. System Description

### A. Mapping of the Distributed-TDM System to a MIMO System

In order to facilitate the derivation of a general solution, the above presented system is mapped to an NxN MIMO system, as illustrated in Fig. 3. Note that N is the number of TDM sub-channels that equals the number of ADCs and DACs. As shown earlier, the TDM sub-channels obtained from *distributed-TDM* may have considerably different characteristics. A TDM block consists of symbols $x_1(n)$ ... $x_N(n)$ that are transmitted sequentially in time via transmitters $T_1(z)$ ... $T_N(z)$. Receivers $R_1(z)$ ... $R_N(z)$ output the samples of the TDM block $y_1(n)$ ... $y_N(n)$.

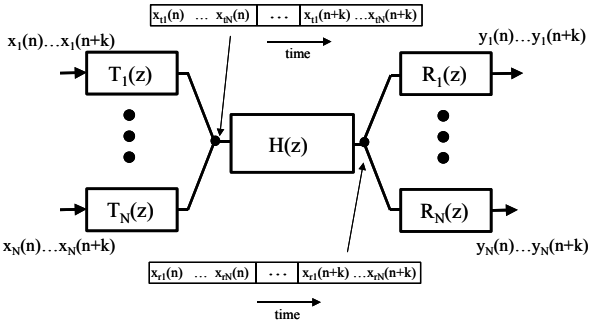


Fig. 3. NxN *distributed-TDM* system.

To illustrate the concept, a 2x2 case is shown in Fig. 4. The transmitter-channel-receiver MIMO model is characterized by the filters $\underline{\mathbf{p}}_{11}, \underline{\mathbf{p}}_{12}, \underline{\mathbf{p}}_{21}, \underline{\mathbf{p}}_{22}$. Assuming that a pulse is transmitted on the first input at time *n*, the noisy measured responses are $\underline{\tilde{\mathbf{p}}}_{11}$ at the first output and $\underline{\tilde{\mathbf{p}}}_{21}$ at the second output. The samples transmitted/received by a particular transmitter/receiver are shown in black (black dots) while the ones that are "skipped" are shown in grey (white dots).

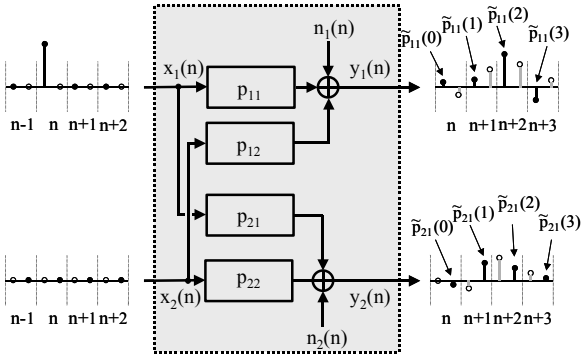


Fig. 4. 2x2 multi-channel system with example channel response.

Another way to view this mapping is that the samples of each TDM block are considered as a vector. This has the advantage of transforming a Single-Input-Single-Output cyclostationary channel into a MIMO time-invariant channel.

### B. Noise Analysis

Prior to deriving the MIMO pre-emphasis filter, an analysis of the noise sources is needed to properly formulate the optimization problem. In the following, the definition of noise is much broader than the typically assumed AWGN thermal noise.

Noise that occurs directly in the *voltage domain* is independent of the transmitted data and can be either static (receiver offset, required overdrive), or dynamic (reference level noise and thermal noise from receiver circuitry). *Time domain* noise, static (phase offset) or dynamic (jitter), while independent of the transmitted data in the time domain, becomes dependent on the data when "translated" to the voltage domain.

Dynamic voltage domain noise can as usual be treated as AWGN. The effect of static voltage domain noise is taken into account in probability of error calculations by subtracting it from the minimum distance between the points of the received constellation.

Time domain noise can be mapped to the voltage domain using the quadratic approximation of a sine-wave [5], which considers the worst case mapping and ignores the dependency on the data:

$$V_{noise} = \frac{\Delta V_r}{2}\left(1-\cos\left(\pi\frac{t_{error}}{T_{symbol}}\right)\right) \le \frac{\Delta V_r}{4}\left(\pi\frac{t_{error}}{T_{symbol}}\right)^2 \qquad (1.1)$$

In the above equation, $\Delta V_r$ is the amplitude of the main harmonic of the receiver signal, $t_{error}$ is the sampling timing error, and $T_{symbol}$ is the symbol period. The above approximation is quite accurate for the low-pass channels that characterize high-speed links. Thus, assuming that the jitter noise is white with a Gaussian distribution in the time domain, its distribution in the voltage domain becomes chi-square with one degree of freedom. The variance of the jitter-induced noise in the voltage domain is bound as shown below:

$$E\left(V_{noise}^2\right) \le 3\Delta V_r^2\left(\frac{\pi}{2T_{symbol}}\right)^4\left(E\left(t_{error}^2\right)\right)^2 \qquad (1.2)$$

where $t_{error} \sim N\left(0,\sigma_{jitter}^2\right)$ and the chi-square distribution is approximated by a Gaussian. Static timing noise can be accounted for by subtracting it from $T_{symbol}$ in (1.1).

To summarize, voltage domain noise is signal-independent and can be traded for ISI. Time domain noise is signal-dependent when mapped to voltage domain and cannot be traded for ISI, although it severely affects the performance of the system. Static noise decreases time and voltage margins and is taken as a constant correction factor, while dynamic noise, both in time and voltage domain is assumed to be Gaussian and treated statistically.

## III. EQUALIZATION

### A. Problem Formulation

The link described in Section II can be modeled as the MIMO system shown in Fig. 5, which includes the transmit pre-emphasis filter, the channel, and receiver scaling.

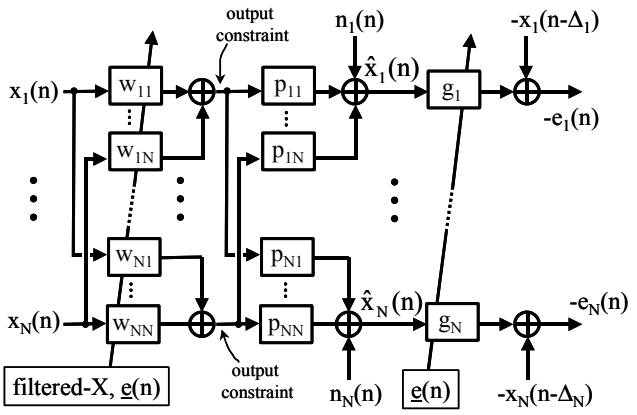


Fig. 5. NxN MIMO system with transmit pre-emphasis filter and receiver scaling.

An important constraint in the design of the transmit pre-emphasis filter is the range of the transmitter output signal. A transmitter precoding solution with an average power constraint was given in [6]. However, the multi-GSa/s high-speed link system requires a probability of error smaller than $10^{-20}$, which implies that clipping noise cannot be tolerated. This leads to the adoption of a peak transmit power constraint, instead of the usual average power constraint. In addition to this, adaptive equalization requires a linear channel and clipping at the transmitter output introduces a non-linear error that causes instability in the adaptation procedure.

The formulation used in this paper is based on the NxN case of a system presented in [7]. A Minimum Mean-Square Error (MMSE) criterion is initially employed, and filters are assumed to have finite length. The response of the system from input *k* to output *l* at sample time *n* is given by:

$$\hat{x}_{lk}(n) = \underline{\mathbf{x}}_k^{\mathbf{T}}(n)\left(\mathbf{P}_{l1}\underline{\mathbf{w}}_{1k} + \ldots + \mathbf{P}_{lj}\underline{\mathbf{w}}_{jk} + \ldots + \mathbf{P}_{lN}\underline{\mathbf{w}}_{Nk}\right) \tag{2}$$

where $\underline{\mathbf{x}}_k(n) = \left[x_k(n) \quad \ldots \quad x_k(n-L-v+1)\right]^{\mathbf{T}}_{\mathbf{(L+v)\times 1}}$ is the data vector at input *k*, $\underline{\mathbf{w}}_{jk}(n) = \left[w_{jk}(0) \quad \ldots \quad w_{jk}(L-1)\right]^{\mathbf{T}}_{\mathbf{L\times 1}}$ is the pre-emphasis filter from data input *k* to transmitter output *j*, and the channel convolution matrix $\mathbf{P}_{lj}$ is defined as:

$$\mathbf{P}_{lj} = \begin{bmatrix} p_{lj}(0) & \ldots & p_{lj}(v) & 0 & 0 & 0 \\ 0 & p_{lj}(0) & \ldots & p_{lj}(v) & 0 & 0 \\ 0 & 0 & \ldots & \ldots & \ldots & 0 \\ 0 & 0 & 0 & p_{lj}(0) & \ldots & p_{lj}(v) \end{bmatrix}^{\mathbf{T}}_{\mathbf{(L+v)\times L}} \tag{3}$$

where the maximum ISI spread (measured in number of TDM blocks) for all $\mathbf{P}_{lj}$ is ν, and the pre-emphasis filter length is L. The system is fully described by:

$$\underline{\hat{\mathbf{x}}}(n)_{\mathbf{N\times 1}} = \mathbf{X}(n)_{\mathbf{N\times N^2(L+v)}}\,\boldsymbol{\Psi}_{\mathbf{N^2(L+v)\times N^2L}}\,\underline{\mathbf{W}}_{\mathbf{N^2L\times 1}} + \underline{\mathbf{N}}(n)_{\mathbf{N\times 1}} \tag{4}$$

where the channel matrix $\boldsymbol{\Psi}$ is defined as:

$$\boldsymbol{\Psi} = \begin{bmatrix} \mathbf{P} & & \mathbf{0} \\ & \ldots & \\ \mathbf{0} & & \mathbf{P} \end{bmatrix}_{\mathbf{N^2(L+v)\times N^2L}} \qquad \mathbf{P} = \left[\mathbf{P}_1^{\mathbf{T}} \ldots \mathbf{P}_N^{\mathbf{T}}\right]^{\mathbf{T}}_{\mathbf{N(L+v)\times NL}} \tag{5}$$

with $\mathbf{P}_k = \left[\mathbf{P}_{k1} \cdots \mathbf{P}_{kN}\right], \forall k = 1 \ldots N$, and the pre-emphasis filter $\underline{\mathbf{W}}$ is defined as:

$$\underline{\mathbf{W}} = \left[\underline{\mathbf{w}}_1^{\mathbf{T}} \quad \ldots \quad \underline{\mathbf{w}}_N^{\mathbf{T}}\right]^{\mathbf{T}}_{\mathbf{N^2L\times 1}}, \; \underline{\mathbf{w}}_k = \left[\underline{\mathbf{w}}_{1k}^{\mathbf{T}} \quad \ldots \quad \underline{\mathbf{w}}_{Nk}^{\mathbf{T}}\right]^{\mathbf{T}}_{\mathbf{NL\times 1}} \tag{6}$$

and the input data matrix $\mathbf{X}(n)$ is defined as:

$$\mathbf{X}(n) = \left[\mathbf{X}_1(n) \quad \ldots \quad \mathbf{X}_N(n)\right]_{\mathbf{N\times N^2(L+v)}} \tag{7}$$

$$\mathbf{X}_k(n) = \begin{bmatrix} \underline{\mathbf{x}}_k^{\mathbf{T}}(n) & & \underline{\mathbf{0}} \\ & \ldots & \\ \underline{\mathbf{0}} & & \underline{\mathbf{x}}_k^{\mathbf{T}}(n) \end{bmatrix}_{\mathbf{N\times N(L+v)}}.$$

The noise vector $\underline{\mathbf{N}}(n)$ represents static and dynamic voltage domain noise as well as static and dynamic time domain noise mapped to the voltage domain**.** The *unbiased error* is defined by:

$$\underline{\mathbf{e}}(n)_{\mathbf{N\times 1}} = \frac{1}{\sqrt{\overline{E}_x}}\left(\underline{\mathbf{x}}(n-\underline{\boldsymbol{\Delta}})_{\mathbf{N\times 1}} - \mathbf{g}_{\mathbf{N\times N}}\,\underline{\hat{\mathbf{x}}}(n)_{\mathbf{N\times 1}}\right) \tag{8}$$

where $\mathbf{g}$ is an NxN diagonal matrix with diagonal elements equal to $g_i, \forall i = 1 \ldots N$, where $g_i$ is the scaling required at receiver *i*, so that the decisions include no bias. $\overline{E}_x$ is the average symbol energy, the vector of *delayed* transmitted data is $\underline{\mathbf{x}}(n-\underline{\boldsymbol{\Delta}}) = \left[x_1(n-\Delta_1) \quad \ldots \quad x_N(n-\Delta_N)\right]^{\mathbf{T}}_{\mathbf{1\times N}}$, and $\underline{\boldsymbol{\Delta}}$ is a vector containing the decision delays of the corresponding inputs.

The total mean square error (9) represents the sum of the inverse of Signal-to-Noise Ratios (SNRs) and hence directly reflects the performance of the system.

$$\xi = E\left(\underline{\mathbf{e}}^{\mathbf{T}}(n)_{\mathbf{1\times N}}\,\underline{\mathbf{e}}(n)_{\mathbf{N\times 1}}\right) \tag{9}$$

At this point, it is useful to observe the duality between receiver equalization [8] and transmit pre-emphasis. While receiver equalization attempts to flatten the frequency response with an effective gain of one, transmit pre-emphasis with an output range constraint attempts to flatten the frequency response at a level below the biggest channel attenuation, [3]. Similarly to the well-known noise enhancement problem related to receiver equalization, the noise at the receiver is relatively amplified due to the signal attenuation.

Retaining only the part of noise that is independent of the data signal, having variance $\sigma^2_{total}$, (9) is expanded to:

$$\xi = N - 2\underline{\mathbf{W}}^{\mathbf{T}}\boldsymbol{\Psi}^{\mathbf{T}}\mathbf{G}^{\mathbf{T}}\underline{\mathbf{I}}_{\boldsymbol{\Delta}}^{\mathbf{T}} + \underline{\mathbf{W}}^{\mathbf{T}}\boldsymbol{\Psi}^{\mathbf{T}}\mathbf{G}^{\mathbf{T}}\,\mathbf{G}\boldsymbol{\Psi}\underline{\mathbf{W}} + \frac{tr\{\mathbf{g}^2\}}{SNR_0} \tag{10}$$

where the scaling matrix G is defined through the expression $\mathbf{X}^{\mathbf{T}}(n)\,\mathbf{g}^{\mathbf{T}} = \mathbf{G}^{\mathbf{T}}\,\mathbf{X}^{\mathbf{T}}(\mathbf{n})$ and $SNR_0 = \overline{E}_x / \sigma^2_{total}$. Also, the vector $\underline{\mathbf{I}}_{\boldsymbol{\Delta}} = \left[\underline{\mathbf{1}}_{\Delta_1} \quad \underline{\mathbf{0}} \quad \ldots \quad \underline{\mathbf{0}} \quad \underline{\mathbf{0}} \quad \underline{\mathbf{1}}_{\Delta_2} \quad \ldots \quad \underline{\mathbf{0}} \quad \ldots \quad \underline{\mathbf{0}} \quad \ldots \quad \underline{\mathbf{0}} \quad \underline{\mathbf{1}}_{\Delta_N}\right]_{\mathbf{1\times N^2(L+v)}}$

represents the system delay, where the vector $\underline{\mathbf{1}}_{\Delta_l} = [0 \quad \ldots \quad 0 \quad 1 \quad 0 \quad \ldots \quad 0]_{\mathbf{1\times(L+v)}}$ has a 1 at position $\Delta_l + 1$.

It may seem that an alternative approach to obtaining a solution to this problem is to determine the well-known receiver equalizer, and subsequently place it at the transmitter (possibly combined with proper scaling). However, such a method ignores the basic fact that the channel and equalizer matrices do not in general commute. If the equalizer is equal to the inverse of the channel (perfect zero-forcing), then under some assumptions the matrices may commute. Finite-length implementations cannot generally achieve perfect zero-forcing. Moreover, MMSE Linear Equalizer (MMSE-LE) solutions have been shown to be more desirable than Zero-Forcing Linear-Equalizer (ZF-LE) in circumstances of high interference.

### *B. Optimal Solution*

The equalizer design may be formulated as an optimization problem, where the objective is the minimization of (10). This problem is subject to two sets of constraints. First, the receiver scaling must be such that the decisions are unbiased. Therefore, the following must hold:

$$g_k \underline{\mathbf{w}}_k^{\mathbf{T}} \mathbf{P}_k^{\mathbf{T}} \underline{\mathbf{1}}_{\Delta_k}^{T} = 1, \forall k = 1 \ldots N \tag{11}$$

which implies that:

$$\underline{\mathbf{W}}^{\mathbf{T}} \boldsymbol{\Psi}^{\mathbf{T}} \mathbf{G}^{\mathbf{T}} \underline{\mathbf{I}}_{\Delta}^{\mathbf{T}} = N \,. \tag{12}$$

Then, (10) becomes:

$$\xi = -N + \sum_{k=1}^{N} \underline{\mathbf{w}}_k^{\mathbf{T}} \left( \sum_{l=1}^{N} \frac{\mathbf{P}_l^{\mathbf{T}} \mathbf{P}_l}{\left( \underline{\mathbf{1}}_{\Delta_l} \mathbf{P}_l \underline{\mathbf{w}}_l \right)^2} \right) \underline{\mathbf{w}}_k + \frac{1}{SNR_0} \sum_{k=1}^{N} \left( \underline{\mathbf{1}}_{\Delta_k} \mathbf{P}_k \underline{\mathbf{w}}_k \right)^{-2} . \tag{13}$$

Additionally, the peak-transmitted signal on each of the outputs must be constrained:

$$\tilde{h}_j(\underline{\mathbf{W}}) = \sum_{k=1}^{N} \sum_{i=0}^{L-1} \left| w_{jk}(i) \right| \le 1, \ \forall j = 1 \ldots N \,. \tag{14}$$

The filter design is expressed as an optimization problem involving the minimization of (13) subject to the constraints of (14). It appears that optimal closed-form solutions are impossible to obtain. Thus, the question is whether numerical solutions can be efficiently produced. It is known that if the optimization problem is *convex*, then there exist several algorithms (such as interior point methods) for obtaining its solution, [9].

The optimization problem is convex if (a) the constraints form a convex set, and (b) the objective function is convex. The constraints can be viewed as limits on the sums of the $l_1$ norms of the filter vectors, hence they do form a convex set. However, the objective function is not convex. The Appendix gives a proof by providing a simple counter-example.

The consequence of non-convexity is that numerical techniques involving gradient methods do not necessarily yield a globally optimal solution. In the following, sub-optimal solutions are investigated, which still result in very good system performance.

### *C. Sub-optimal Solutions*

Two sub-optimal approaches are presented that are based on the above formulation and work well at moderate to high SNR. These allow the implementation of simple LMS-type adaptive algorithms, which are proposed in the next section.

According to the first approach, which does not have a closed form solution, the inequality constraints in (14) are substituted by equalities. This is equivalent to forcing the maximum output of every pre-emphasis filter to be equal to the maximum transmitter range, thus transmitting the maximum available power into the channel and putting more strain on the pre-emphasis.

With the second approach, one first finds the unconstrained ZF-LE (ZFEU) solution, and then scales all the transmit filters by the same amount obtained from the unconstrained pre-emphasis filter with largest peak output, as shown below:

$$\underline{\mathbf{W}} = \frac{\underline{\mathbf{W}}_{ZFEU}}{\max_{j=1\ldots N}(\tilde{h}_j(\underline{\mathbf{W}}_{ZFEU}))}, \ \underline{\mathbf{W}}_{ZFEU} = \left( \boldsymbol{\Psi}^{\mathbf{T}} \boldsymbol{\Psi} \right)^{-1} \boldsymbol{\Psi}^{\mathbf{T}} \mathbf{G}^{\mathbf{T}} \underline{\mathbf{I}}_{\boldsymbol{\Delta}}^{\mathbf{T}} . \tag{15.1-2}$$

The optimal delay $\underline{\mathbf{I}}_{\boldsymbol{\Delta}}$ vector in (15.1-2) is determined from the set of delays $\underline{\Delta} = [\Delta_1 \quad \ldots \quad \Delta_N]$ that result in minimum total square error on each channel (15.3).

$$\Delta_j = \arg\max \left( diag_{j(L+v):(j+1)(L+v)-1} \left( \boldsymbol{\Psi} \left( \boldsymbol{\Psi}^{\mathbf{T}} \boldsymbol{\Psi} \right)^{-1} \boldsymbol{\Psi}^{\mathbf{T}} \right) \right) \tag{15.3}$$

### *D. Adaptive Solution*

While closed form solutions have theoretical value, their mathematical complexity makes them impractical for the implementation of a real high-speed system. However, both sub-optimal approaches, individual scaling and maximal scaling, (15.1-3), can be implemented adaptively in a very simple manner using a modification of the multi-channel, multiple-error filtered-X LMS algorithm [10] as shown in Fig. 5. The pre-emphasis filter tap adaptation is described below:

$$SE_n = \underline{\mathbf{e}}(n)^{\mathbf{T}} \underline{\mathbf{e}}(n) \tag{16.1}$$

$$\hat{\nabla}_n(\underline{\mathbf{W}}) = \frac{\partial SE_n}{\partial \underline{\mathbf{W}}} = -\frac{2}{\sqrt{E_x}} \underline{\mathbf{U}}(n)^{\mathbf{T}} \mathbf{g}^{\mathbf{T}} \underline{\mathbf{e}}(n), \ \underline{\mathbf{U}}(n) = \mathbf{X}(n) \boldsymbol{\Psi} \tag{16.2-3}$$

$$\underline{\tilde{\mathbf{W}}}_{n+1} = \underline{\mathbf{W}}_n - \mu_w \hat{\nabla}_n(\underline{\mathbf{W}}) = \underline{\mathbf{W}}_n + \frac{2\mu_w}{\sqrt{E_x}} \underline{\mathbf{U}}(n)^{\mathbf{T}} \mathbf{g}^{\mathbf{T}} \underline{\mathbf{e}}(n) \tag{16.4}$$

$$\underline{\mathbf{W}}_{n+1} = \left[ \frac{\underline{\tilde{\mathbf{w}}}_{11}^{\mathbf{T}}(n+1)}{\tilde{h}_1(\underline{\tilde{\mathbf{W}}}_{n+1})_{\mathbf{1\times L}}} \quad \ldots \quad \frac{\underline{\tilde{\mathbf{w}}}_{NN}^{\mathbf{T}}(n+1)}{\tilde{h}_N(\underline{\tilde{\mathbf{W}}}_{n+1})_{\mathbf{1\times L}}} \right]^{\mathbf{T}}_{\mathbf{N^2L\times 1}} \tag{16.5}$$

$$\underline{\mathbf{W}}_{n+1} = \frac{1}{\max_{j=1\ldots N}(\tilde{h}_j(\underline{\tilde{\mathbf{W}}}_{n+1}))} \underline{\tilde{\mathbf{W}}}_{n+1} \,. \tag{16.6}$$

In the above, $\underline{\mathbf{U}}(n)$ is a filtered-X signal available in the transmitter. The difference between the two approaches lies in the scaling of the filter coefficients after each iteration, which is shown in (16.5) for individual scaling and in (16.6) for maximal scaling.

The scaling values are adapted as shown below:

$$\hat{\nabla}_n(\mathbf{g}) = \frac{\partial SE_n}{\partial diag(\mathbf{g}_n)} = -\frac{2}{\sqrt{E_x}} diag(\underline{\hat{\mathbf{x}}}(n) \underline{\mathbf{e}}^{\mathbf{T}}(n)) \tag{17.1}$$

$$diag(\mathbf{g}_{n+1}) = diag(\mathbf{g}_n) + \frac{2\mu_g}{\sqrt{E_x}} diag(\underline{\hat{\mathbf{x}}}(n) \underline{\mathbf{e}}^{\mathbf{T}}(n)) \,. \tag{17.2}$$

The scaling loop has to converge much faster than the equalizer loop, since it provides the reference for the equalizer update.

In high-speed links, the algorithm is only intended to run when the configuration of the system has changed, and occasionally to adapt to variations of the transmission environment. Therefore, convergence time is not a huge issue and the implementation of a delayed version of the above algorithm is possible. The delayed version of the algorithm updates the equalizer coefficients only when new error information is available from the receiver on the back-channel.

## IV. Experimental and Simulation Results

This section presents the results of performance simulations with the channel data obtained from the experimental test-bed presented in [3]. The original system employed several algorithms for static noise correction and channel equalization, and operated at 8 GSa/s with 2-PAM and 4-PAM modulation.

The simulation environment takes into account the effect of noise sources described in Section II.B. and obtains the coefficients of the pre-emphasis filters using the adaptive algorithms in Section III.D. The setup uses 4-PAM modulation and the peak transmitter output range is 750mV, with L=1 to 4 taps per pre-emphasis filter and $\nu$=4.

To illustrate the channel-to-channel variations, pulse responses from all eight channels obtained in [3] are shown in Fig. 6. Clearly, not only are the ISI "profiles" distinct, but also the channel attenuation differs.

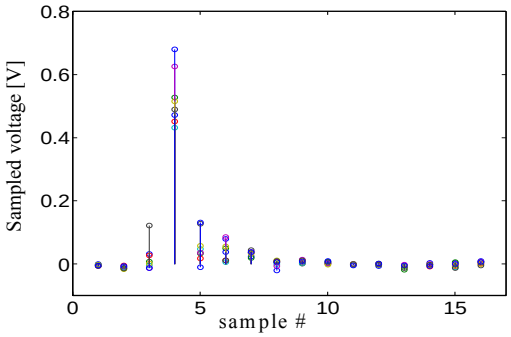


Fig. 6. Overlaid pulse responses of the 8 TDM sub-channels.

Learning curves of total cost function $\xi$ are shown in Fig. 7 for both individual scaling per channel and maximal scaling per channel, together with the cost function of scaled ZFEU with $\sigma_{total}$=4mV, $\sigma_{jitter}$=6ps, L=4.

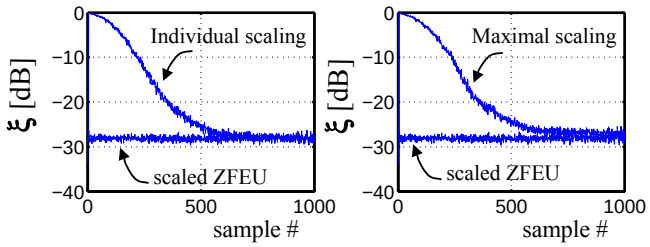


Fig. 7. Cost function learning curves.

Evidently, individual scaling performs slightly better than maximal. Adaptive maximal scaling converges to the scaled ZFEU solution with a misadjustment penalty, which depends on the convergence rate and the number of filter taps. The advantage of individual over maximal scaling is highly dependent on the type of the channel. In certain cases, the benefit of individual scaling is offset by the increase in residual ISI.

Learning curves for cost and scaling per channel are shown in Fig. 8. Notice that the scaling curves are a lot smoother and converge faster than the cost function. This is characteristic of the filtered-X algorithm, since the filtered-X value contains no noise information and hence is less robust to noise than the scaling loop at the receiver.

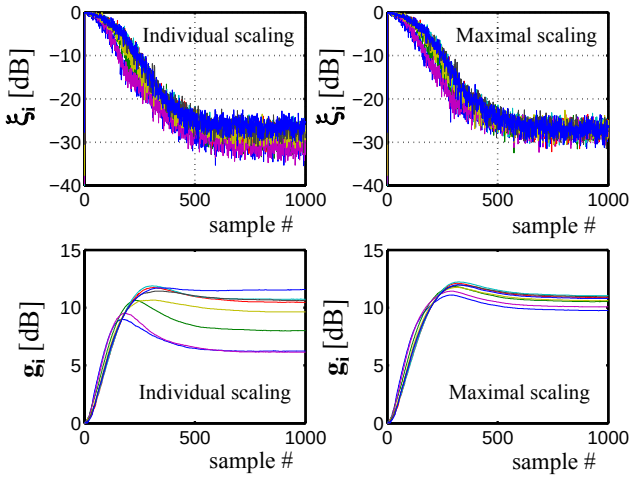


Fig. 8. Cost and scaling learning curves per channel.

The performance of the system is illustrated in Fig. 9, for different L and noise values. Voltage domain noise is composed of thermal noise and voltage reference noise with noise power $\sigma_{total}$. Time domain noise consists of the clock jitter having standard deviation $\sigma_{jitter}$. After the taps are obtained from the simulation, the probability of error is estimated from the SNR value corresponding to the tap coefficients.

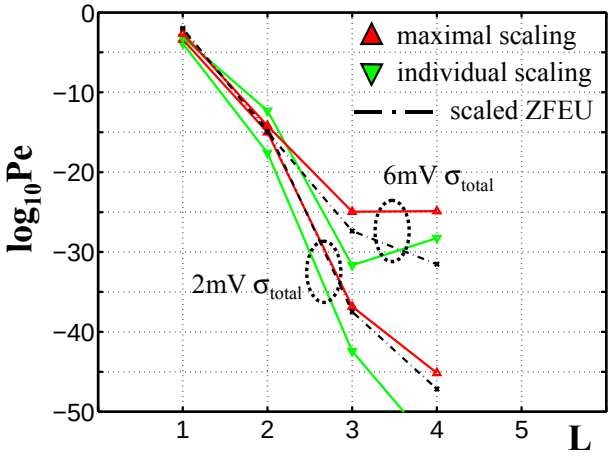


Fig. 9. Pe versus filter length, $\sigma_{jitter}$=6ps.

Different jitter noise values do not affect the filter coefficients, since these are averaged over multiple runs. Jitter noise, however, limits the performance of the system since it cannot be traded for ISI. At high noise setting, the point of noise-ISI tradeoff occurs at smaller number of taps (L=3-4), than at lower noise setting, indicating that the algorithms are trading noise for ISI.

Both individual scaling and maximal scaling have similar performance on the presented type of system, with some advantage obtained with individual scaling. However, both algorithms offer similar performance to closed form solutions with significantly less computation. The residual non-averaged jitter noise limits the accuracy of the probability of error calculation method, especially for very low probability of error values that occur when a large number of taps is used. Another effect that was observed in the results is the non-monotonicity of the probability of error curve with the number of taps. This is attributed to symbol spaced equalization and jitter. Symbol spaced equalization does not have direct control over the width of the data eye. Hence, although more taps in the filter mean less residual ISI, wider data eye is not guaranteed. It is possible in cases with dominant jitter noise to improve voltage margins with more taps, but degrade timing margins at the same time, and thus degrade the probability of error.

## V. Conclusion

Very high-speed serial links are starting to use multi-level modulation and TDM in order to avoid technology limitations in transceiver and transmission medium design. *Distributed-TDM* has been proposed as a potential remedy to the bandwidth penalty associated with standard TDM. This paper proposes the methodology of mapping the *distributed-TDM* system to a MIMO system with transmit pre-emphasis. This MIMO transmit pre-emphasis formulation has previously been proposed for channel inversion in multi-channel sound reproduction systems. Very low bit error rate requirements in the absence of coding impose a peak transmit power constraint instead of the traditionally used average power constraint. The resulting optimization problem is shown to be non-convex, making optimal solutions very difficult to obtain. Sub-optimal solutions based on equality-constrained optimization and scaled unconstrained zero-forcing linear equalizer are presented. Targeting the practical implementation of the system, adaptive solutions are derived based on the modification of the multi-channel, multi-error, filtered-X LMS algorithm. This modification entails simultaneous adaptation of the pre-emphasis filter coefficients at the transmitter side and of the scaling coefficients at the receiver side. In order to avoid the introduction of non-linear clipping noise, the transmit filters are either scaled independently or the same scaling is applied across all of them.

## APPENDIX

Consider a system where $N=2$, $L=1$ and $\nu=0$, implying that there is no ISI, and the delay vector becomes irrelevant. Then, the objective function (13) becomes:

$$\xi=\frac{1}{\left(\mathbf{P}_1\underline{\mathbf{w}}_1\right)^2}\left[\frac{1}{SNR_o}+\left(\mathbf{P}_1\underline{\mathbf{w}}_2\right)^2\right]+\frac{1}{\left(\mathbf{P}_2\underline{\mathbf{w}}_2\right)^2}\left[\frac{1}{SNR_o}+\left(\mathbf{P}_2\underline{\mathbf{w}}_1\right)^2\right] \quad \text{(A.1)}$$

A function is convex only if it is convex on all lines. Let:

$$\left[\underline{\mathbf{w}}_1 \quad \underline{\mathbf{w}}_2\right]=\left[\underline{\mathbf{w}}_{o,1} \quad \underline{\mathbf{w}}_{o,2}\right]+t\cdot\left[\underline{\mathbf{w}}_{s,1} \quad \underline{\mathbf{w}}_{s,2}\right]. \quad \text{(A.2)}$$

Therefore, $\xi$ is convex in $\left[\underline{\mathbf{w}}_1 \quad \underline{\mathbf{w}}_2\right]$ only if it is convex in $t$ for all $\left[\underline{\mathbf{w}}_{o,1} \quad \underline{\mathbf{w}}_{o,2}\right]$ and $\left[\underline{\mathbf{w}}_{s,1} \quad \underline{\mathbf{w}}_{s,2}\right]$. By substituting (A.2) in (A.1), it becomes clear that each of the two terms of $\xi$ is a ratio of quadratic expressions in $t$, which are known to be non-convex. Therefore, (A.1) is not convex in general.


## Acknowledgment

The authors would like to acknowledge conversations with C.-K. K. Yang and W. Ellersick, during the work on [3], that uncovered the problems addressed in this paper. V. Stojanovic also thanks I. Stojanovic for full-hearted help in development of the paper and invaluable technical discussions.


## References


[1] R. Farjad-Rad, C.-K. K. Yang and M. A. Horowitz, "A 0.3-μm CMOS 8-GS/s 4-PAM Serial Link Transceiver," *IEEE Journal of Solid-State Circuits,* vol 35, no 5, May 2000, pp. 757-64.

[2] J. Zerbe *et al*, "A 2Gb/s/pin 4-PAM Parallel Bus Interfaces with Transmit Crosstalk Cancellation Equalization and Integrating Receivers," *IEEE ISSCC Dig. of Tech. Papers*, Feb. 2001, San Francisco, pp. 66-7, 432.

[3] W. Ellersick, C.-K. K. Yang, V. Stojanovic, S. Modjtahedi and M. A. Horowitz, "A serial-link transceiver based on 8 Gsample/s A/D and D/A converters in 0.25um CMOS," *IEEE ISSCC Dig. of Tech. Papers*, Feb. 2001, San Francisco, pp. 58-9, 430.

[4] E. Ginzton, W. Hewlett, J. H. Jasberg and J. D. Noe, "Distributed Amplification," *Proceedings of the IRE*, vol. 36, August 1948, pp. 956-69.

[5] W. Ellersick, *Data Converters for High Speed CMOS Links, Ph.D. thesis*, Stanford University, August 2001.

[6] B. R. Vojčić and W. M. Jang, "Transmitter Precoding in Synchronous Multiuser Communications," *IEEE Transactions on Communications*, vol. 46, no. 10, October 1998, pp. 1346-55.

[7] P. A. Nelson, F. O-Bustamante and H. Hamada, "Inverse Filter Design and Equalization Zones in Multichannel Sound Reproduction," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 3, May 1995, pp.185-92.

[8] J. Proakis, and M. Salehi, *Communication Systems Engineering,* Prentice Hall 1994.

[9] S. Boyd and L. Vandenberghe, *Lecture Notes for Convex Optimization*, Stanford University, Winter 2000-01, (http://www.stanford.edu/class/ee364/).

[10] P. A. Nelson, H. Hamada and S. J. Elliot, "Adaptive Inverse Filters for Stereophonic Sound Reproduction," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, July 1992, pp. 1621-32.