

Transmitter and Receiver Equalizers Optimization Methodologies for High-Speed Links in Industrial Computer Platforms Post-Silicon Validation

Francisco E. Rangel-Patiño^{1,2}, José E. Rayas-Sánchez¹, and Nagib Hakim²

¹Department of Electronics, Systems, and Informatics, ITESO – The Jesuit University of Guadalajara
Tlaquepaque, Jalisco, 45604 Mexico

²Intel Corp.

e-mail: francisco.rangel@intel.com

Abstract — As microprocessor design scales to nanometric technology, traditional post-silicon validation techniques are inappropriate to get a full system functional coverage. Physical complexity and extreme technology process variations introduce design challenges to guarantee performance over process, voltage, and temperature conditions. In addition, there is an increasingly higher number of mixed-signal circuits within microprocessors. Many of them correspond to high-speed input/output (HSIO) links. Improvements in signaling methods, circuits, and process technology have allowed HSIO data rates to scale beyond 10 Gb/s, where undesired effects can create multiple signal integrity problems. With all of these elements, post-silicon validation of HSIO links is tough and time-consuming. One of the major challenges in electrical validation of HSIO links lies in the physical layer (PHY) tuning process, where equalization techniques are used to cancel these undesired effects. Typical current industrial practices for PHY tuning require massive lab measurements, since they are based on exhaustive enumeration methods. In this paper, direct and surrogate-based optimization methods, including space mapping, are proposed based on suitable objective functions to efficiently tune the transmitter and receiver equalizers. The proposed methodologies are evaluated by lab measurements on realistic industrial post-silicon validation platforms, confirming dramatic speed up in PHY tuning and substantial performance improvement.

Index Terms — ANN, Broyden, channel, crosstalk, CTLE, DoE, equalization, Ethernet, eye diagram, FIR, HSIO, ISI, jitter, Kriging, metamodels, optimization, PCIe, post-silicon validation, receiver, SATA, SFP, signal integrity, space mapping, surrogates, system margining, transmitter, tuning, USB.

I. INTRODUCTION

Technology scaling and advanced silicon packaging techniques are allowing high density integration. However, as process technologies scale down to nanometric dimensions, traditional IC design methods are challenged by the problem of increased silicon process variation. The combined effects of increased product complexity, performance requirements, and time-to-market (TTM) commitments have added tremendous pressure on post-silicon validation [1].

A significant portion of the circuits to be validated in modern microprocessors corresponds to high-speed input/output (HSIO) links. Undesired effects such as jitter, inter-symbol interference (ISI), crosstalk and others, can create multiple signal integrity problems in HSIO circuits, making

maximum bus speeds difficult to achieve in practice. This problem is aggravated by the fact that channel speeds keep increasing from one generation bus technology to the next one. This is of particular concern for HSIO interfaces, such as Peripheral Component Interconnect Express (PCIe), Serial Advanced Technology Attachment (SATA), Universal Serial Bus (USB), and Ethernet interfaces.

Training algorithms and physical layer (PHY) tuning are two important components in modern HSIO links. PHY tuning knobs are usually embedded in the I/O links, and can be digitally tuned to appropriate values. Considering the large die-to-die process variations, as well as the typical fluctuations in operating conditions, board impedance, channel loss, and different add-in cards/DIMMs, the performance of HSIO links can exhibit large variation. PHY tuning provides a way to reconfigure I/O links to cancel various fluctuations. However, it is usually unknown in pre-silicon which configuration gives the overall best performance, becoming necessary to search for “optimal” PHY tuning knob configurations. Most current industrial practices to perform PHY tuning consist of exhaustive enumeration methods, turning them into the most time-consuming processes in post-silicon validation [2].

This paper presents several optimization techniques based on novel objective functions to optimize the transmitter (Tx) and receiver (Rx) equalizers in a server post-silicon validation platform. This paper essentially summarizes [2], [11] and [25].

The rest of the paper is organized as follows. Section II presents an overview on PHY tuning. Section III describes a holistic optimization approach that merges system margining and jitter tolerance measurements to optimize the Rx equalizer. An efficient optimization methodology is proposed in Section IV to find out the optimal coefficients for a reconfigurable finite impulse response (FIR) filter used on a Tx Ethernet interface. In Section V, a new optimization methodology is proposed to find optimal coefficients for the Tx and Rx in a PCIe equalization (EQ) process. In Section VI, a neural modeling approach is described to efficiently simulate the silicon equalizer Rx. In Section VII, the Broyden-based input space mapping algorithm is exploited to optimize the PHY tuning Rx equalizer. Finally, in Section VIII are discussed the overall results of the proposed techniques.

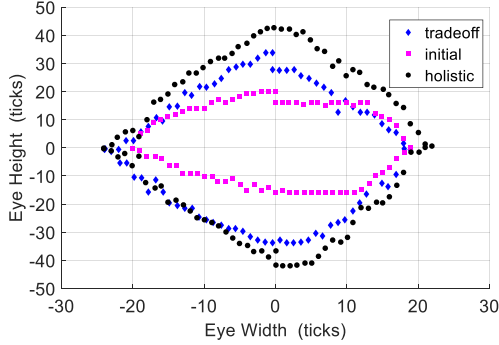


Fig. 3. Eye width vs eye height: comparing proposed methodology against the initial design and the trade-off approach. From [2].

The holistic approach is realized by adding a JTOL penalty function to (3), such that we find EQ knobs settings that optimize the functional eye diagram and simultaneously satisfies the JTOL specified mask. The JTOL system response, $\mathbf{R}_J \in \mathfrak{R}$, consists of measurements of the sinusoidal jitter amplitude,

$$\mathbf{R}_J = \mathbf{R}_J(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta}) = \mathbf{S}_{JA}(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta}) \quad (4)$$

where \mathbf{S}_{JA} is the sinusoidal jitter amplitude. The new optimization problem is then defined as

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} u(\mathbf{x}) \quad \text{subject to } \mathbf{g}(\mathbf{x}) \leq 0 \quad (5)$$

where $\mathbf{g}(\mathbf{x}) = \mathbf{S}_{JA\text{spec}} - \mathbf{S}_{JA}$; $\mathbf{S}_{JA\text{spec}}$ is the JTOL spec mask.

We can define an objective function that covers both the electrical margining system and the JTOL system responses,

$$u(\mathbf{x}) = -w_1 [e_w(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta})] [e_h(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta})] + w_2 [e_{wa}(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta})] + w_3 [e_{ha}(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta})] + r_0^g \|\mathbf{G}(\mathbf{x})\|_2^2 \quad (6)$$

where $r_0^g \in \mathfrak{R}$ is a penalty coefficient and $\mathbf{G}(\mathbf{x})$ is the JTOL penalty vector function defined as,

$$\mathbf{G}(\mathbf{x}) = \max\{\mathbf{0}, \mathbf{g}(\mathbf{x})\} \quad (7)$$

C. Surrogate Model and Optimization

To minimize (6), a surrogate-based optimization strategy is followed. Kriging [12] is selected as the underlying modeling technique, given its adequacy for dealing with multiple optima and non-continuous responses. Our implementation uses the Matlab Kriging toolbox DACE [13]. To enhance the efficiency of our approach, DoE is applied for sampling data.

D. Results

The Rx knobs settings obtained through the optimization process were verified by measuring both the Rx inner eye height/width and jitter tolerance of the PCH. The optimized knobs setting showed an improvement of 175% on eye diagram area as compared to the initial knobs setting, and a 34% improvement as compared to the traditional (tradeoff) approach, as shown in Fig. 3. Similarly, the jitter tolerance results showed a substantial improvement with margins well above the specification limit template, as seen in Fig. 4. The efficiency of this approach was also demonstrated by a

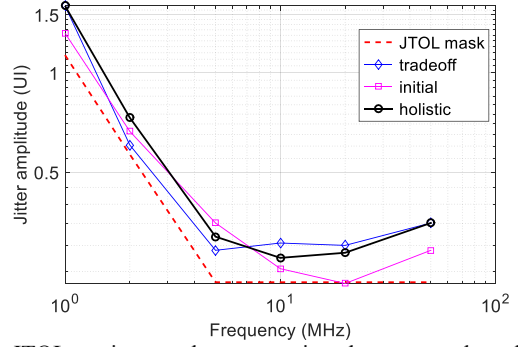


Fig. 4. JTOL testing results: comparing the proposed methodology against the initial design and the trade-off approach. From [2].

significant time reduction on post-silicon validation. While the traditional process requires days for a complete optimization, the method proposed here can be completed in a few hours. The technique can easily be applied to other interfaces such as SATA and PCIe, as demonstrated in [2].

IV. ETHERNET TRANSMITTER EQUALIZATION BY DIRECT OPTIMIZATION

Transceiver modules, such as some Ethernet protocols like the 10-Gigabit Small Form Factor Pluggable (XFP/SFP) and Enhanced SFP (SFP+), are regulated by specifications that ensure consistency between suppliers with requirements for eye mask measurements. These eye mask definitions specify Tx output performance in terms of voltage amplitude and time [14]. Per Ethernet IEEE standard [15], the equalization for SFP+ Tx may be accomplished with a feedforward equalizer (FFE) 3-tap FIR filter. The filter response can be adjusted by controlling the tap number and coefficients values.

Several FIR filter coefficients optimization techniques have been reported [16], [17], [18]; however, all of them are applied at design simulation level. SFP+ Tx FIR filter is not self-adaptive, and then PHY tuning is required during post-silicon validation, being the current practice based on exhaustive enumeration methods.

In [19], we propose a simple yet efficient optimization technique for a reconfigurable FIR filter used in a SFP+ Tx, by defining an effective objective function and by using direct numerical optimization in a post-silicon validation platform.

A. System Test Setup

The test setup is shown in Fig. 5. The eye diagram of the device under test (DUT) is measured at the end of the SFP+ connector using subminiature cables connected to a high-speed, real time oscilloscope. The oscilloscope has enough bandwidth capabilities to measure at least the 5th harmonic of the incoming signal and capabilities for S-parameters de-embedding to eliminate cables insertion loss. A computer executes the algorithm using a fully automated control by accessing the DUT through the test access points registers for the FFE coefficients, sending instruments commands for eye diagram, jitter, and histogram measurements on the scope.

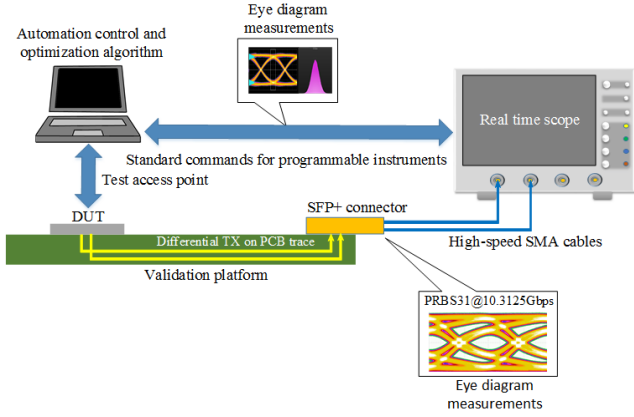


Fig. 5. Test setup for SFP+ Tx optimization. From [19].

B. System Measurements

The definition for eye height is derived from computing the difference between the inner 3σ points on the inside of the histograms of the one and zero levels, as shown in Fig. 6, where σ is the standard deviation of the histograms. The eye width is essentially the effective distance between the inner two 3σ points on the time histograms. To compute jitter, time variances of the rising and falling edges of an eye diagram at the crossing point are captured (see Fig. 6). The time histogram is analyzed to determine the amount of jitter. The peak-to-peak jitter is defined as the full width of the histogram, meaning all data points present.

C. Objective Function Formulation and Optimization

Let $\mathbf{R}_E \in \mathfrak{R}^3$ denote the signal integrity system response, which consists of the eye amplitude histogram mean high $h_{\mu H}$, the histogram mean low $h_{\mu L}$, and the total jitter J_T on the eye diagram,

$$\mathbf{R}_E = \mathbf{R}_E(\mathbf{x}, \boldsymbol{\psi}) = [h_{\mu H}(\mathbf{x}, \boldsymbol{\psi}) \quad h_{\mu L}(\mathbf{x}, \boldsymbol{\psi}) \quad J_T(\mathbf{x}, \boldsymbol{\psi})]^T \quad (8)$$

\mathbf{R}_E is a function of the PHY tuning settings $\mathbf{x} \in \mathfrak{R}^N$ (FIR tap coefficients) and the operating conditions $\boldsymbol{\psi}$. The eye height $e_h \in \mathfrak{R}$ is obtained from

$$e_h(\mathbf{x}, \boldsymbol{\psi}) = h_{\mu H}(\mathbf{x}, \boldsymbol{\psi}) + 3\sigma_H + h_{\mu L}(\mathbf{x}, \boldsymbol{\psi}) - 3\sigma_L \quad (9)$$

where σ_H and σ_L are the standard deviation of the histogram mean high and the histogram mean low, respectively.

Since we want to maximize the eye diagram, our initial objective function consists simply of $-e_h$, however, as the eye width is a function of the total jitter J_T , we must consider J_T in the objective function formulation. e_h and J_T must be scaled by weighting factors $w_1, w_2, \in \mathfrak{R}$ such that they become comparable. Therefore, the objective function is defined as

$$u(\mathbf{x}) = -w_1[e_h(\mathbf{x}, \boldsymbol{\psi})] + w_2[J_T(\mathbf{x}, \boldsymbol{\psi})] \quad (10)$$

The optimization problem for the signal integrity system is

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} u(\mathbf{x}) \quad (11)$$

We now modify the optimization problem such that the optimal set of coefficients maximizes the eye diagram without exceeding the mask limits. The new optimization problem can be defined through a constrained formulation,

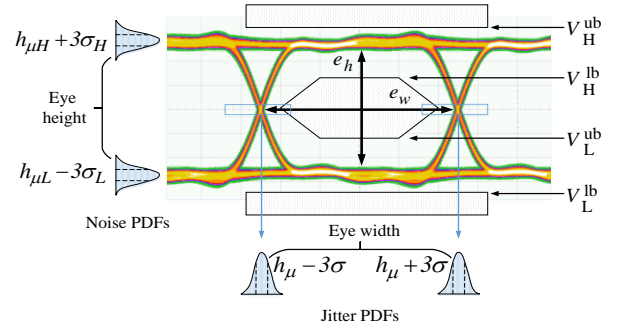


Fig. 6. Eye diagram and mask. From [19].

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} u(\mathbf{x}) \text{ subject to } l_1(\mathbf{x}) \leq 0, l_2(\mathbf{x}) \leq 0 \quad (12)$$

with

$$l_1(\mathbf{x}) = (h_{\mu H} + 3\sigma_H) - V_H^{\text{ub}} \quad (13)$$

$$l_2(\mathbf{x}) = V_L^{\text{lb}} - (h_{\mu L} - 3\sigma_L) \quad (14)$$

where V_H^{ub} and V_L^{lb} are the eye mask specification limits: voltage high upper bound, and voltage low lower bound, respectively. A more convenient unconstrained formulation can be defined by adding a penalty term, as

$$U(\mathbf{x}) = -w_1[e_h(\mathbf{x}, \boldsymbol{\psi})] + w_2[J_T(\mathbf{x}, \boldsymbol{\psi})] + \rho_0^l |L(\mathbf{x})|^2 \quad (15)$$

where $\rho_0^l \in \mathfrak{R}$ is a penalty term and $L(\mathbf{x})$ is the eye mask limit penalty function defined as

$$L(\mathbf{x}) = \max\{0, l_1(\mathbf{x}), l_2(\mathbf{x})\} \quad (16)$$

Our final objective function to optimize eye diagram and meet eye mask specification is

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} U(\mathbf{x}) \quad (17)$$

We find the optimal set of FIR coefficients values \mathbf{x}^* by solving (17) with (15) using the Nelder-Mead method.

D. Results

When the FIR input signal becomes a pseudo-random bit sequence (PRBS) of length $(2^{31} - 1)$ with a 10.3125 Gbps data rate, the resultant eye diagram is shown in Fig. 7. The eye diagram is significantly distorted, with an eye height and eye width of 270 ticks and 189 ticks, respectively. Fig. 7 also shows the zero crossing points on the horizontal axis are not compressed enough, leading to high jitter measurements.

After applying the proposed optimization process, we get optimal Tx EQ coefficients in just 35 iterations. The optimized coefficients substantially improve e_h and e_w , as shown in Fig. 8, being now 864 ticks and 257 ticks, respectively, which corresponds to a 252% improvement as compared to that one with the initial coefficients. The efficiency of this approach was also confirmed by a dramatic time reduction in post-silicon validation, from 4 days in the traditional process based on exhaustive search, to just 2 hours in the proposed method.

V. PCIe TRANSCEIVER EQ BY DIRECT OPTIMIZATION

PCIe is one of the most complex HSIO interfaces [20]. PCIe is a packet based high-speed point-to-point interconnection

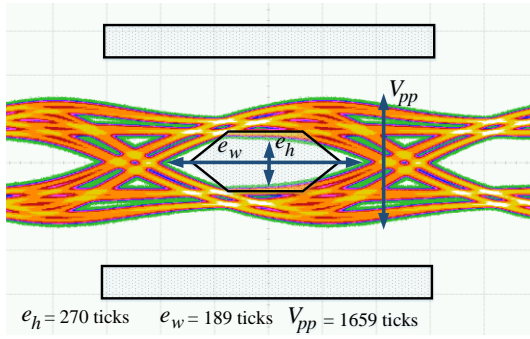


Fig. 7. Eye diagram over mask before optimization. From [19].

technology that evolves with new computer industrial demands [21], and it is the primary interface for a CPU to connect with I/O devices.

The PCIe specification defines an adaptive mechanism for EQ to determine the optimum value of the Tx and Rx EQ coefficients within a fixed time limit. Testing every coefficients combination using an exhaustive enumeration method to find the best one is very time consuming. To speed up this selection, the current practice is to find a subset of coefficient combinations during post-silicon validation, and then program it into the system BIOS. The current industrial method to find the best subset of coefficients consists of using maps of EQ, which are intuitive visual indicators that help experienced post-silicon validation engineers to find the optimal coefficient combination by inspection.

In [22], we propose a simple yet efficient optimization methodology to find the optimal subset of coefficients for the Tx and Rx in a PCIe equalization process, here summarized.

A. Tx and Rx Equalizers

Most Tx serializer-deserializer implementations comprise a FFE 3-tap FIR filter. C_m , C_0 , and C_p represent the three filter taps coefficients. The EQ topology at the Rx may be a combination of a continuous-time linear equalizer (CTLE) that works independently of the clock recovery circuit, and a decision feedback equalizer (DFE). The CTLE is a simple one-tap coefficient (C_r) continuous-time circuit with high-frequency gain boosting, whose transfer function can compensate the channel response [23].

B. Transmitter Equalization Coefficient Matrix

The values of the Tx coefficients are subjected to the following protocol constraints:

$$|C_m| + |C_0| + |C_p| = 1 \text{ subject to } C_0 > 0, C_m \leq 0, C_p \leq 0 \quad (18)$$

These constraints are implemented by determining only C_m and C_p , being C_0 implied by (18). Additionally, the coefficients range and tolerance are constrained by some requirements, as follows.

The coefficients must support all eleven values for the presets, and their respective tolerances, as defined by the Tx preset ratios table in the PCIe specification [21].

In order to keep the output-transmitted power constant with

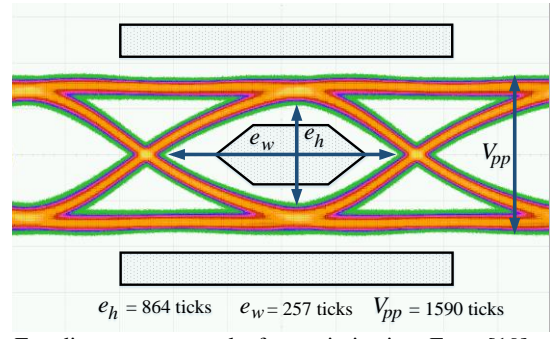


Fig. 8. Eye diagram over mask after optimization. From [19].

respect to coefficients, a full swing (FS) indicates the maximum differential voltage that can be generated by the Tx,

$$FS = |C_m| + |C_0| + |C_p| \quad (19)$$

The flat level voltage should always be greater than the minimum differential voltage that can be generated by the Tx, indicated as the low frequency (LF) parameter,

$$C_0 - |C_m| - |C_p| \geq LF \quad (20)$$

When the above constraints are applied, the resulting coefficients space may be mapped onto a triangular matrix, as shown in Fig. 9, where several EQ maps, one per C_r value, are superimposed. C_m and C_p coefficients are mapped onto the y- and x-axis, respectively. Each matrix cell corresponds to a valid combination of C_m and C_p coefficients, and $u(x^*)$ correspond to a combination of C_m , C_p and C_r that results in an eye diagram qualified as optimum.

Three EQ maps are generated for each of C_r value, and each lane and device pairing may require one or more EQ maps. Current industrial methods, used by experienced validation engineers, consists of visually analyzing each EQ map to select the coefficients C_m and C_p for the FIR filter in the Tx, and C_r for the CTLE in the Rx, that correspond to an eye qualified as optimum. However, this has to be done by ensuring at the same time that the responses around the best C_m - C_p matrix cell are at least 80% the value of that matrix cell (see Fig. 9). Due to the large number of EQ maps, finding the optimal subset of coefficients is usually a very challenging task.

C. Objective Function Formulation and Optimization

We aim at finding the optimal set of coefficients to maximize the functional eye diagram based on the margin response. Here we follow our work in [2] to define the corresponding initial objective function.

As described in previous section, we need to ensure the optimal system margin response is within a suitable area in the coefficients search space of the EQ map. In order to satisfy this requirement, the four margin responses around $u(x^*)$ must be at least 80% of the value of $u(x^*)$, as shown in Fig. 9, where $u_{i,j}$ are the objective function values for the i -th C_m and j -th C_p values, being C_m and C_p the vectors of Tx FIR pre-cursor and post-cursor values, respectively, and C_r is the vector of Rx CTLE coefficient values. This avoids selecting an optimal solution with a too high sensitivity.

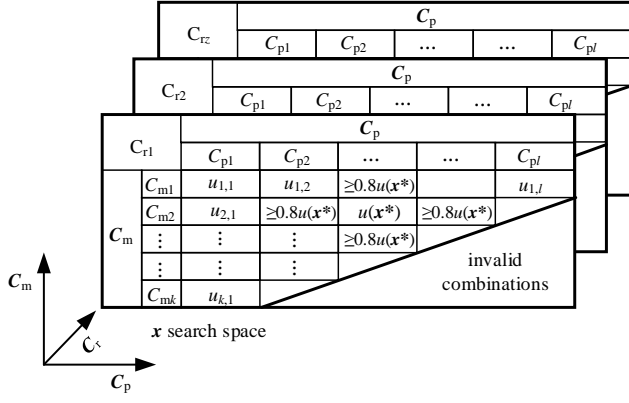


Fig. 9. EQ map coefficients search space for optimization [22].

We now modify the optimization problem such that the optimal set of coefficients maximizes the system margins response without exceeding the limit of $0.8u(x^*)$ in the vicinity. The optimization problem can be defined through a constrained formulation,

$$\begin{aligned} \mathbf{x}^* = \arg \min_{\mathbf{x}} u(\mathbf{x}) \\ \text{subject to } l_1(\mathbf{x}) \leq 0, l_2(\mathbf{x}) \leq 0, l_3(\mathbf{x}) \leq 0, l_4(\mathbf{x}) \leq 0 \end{aligned} \quad (21)$$

with

$$l_1(\mathbf{x}) = u(C_{mi^*+1}, C_{pj^*}, \boldsymbol{\psi}, \boldsymbol{\delta}) - 0.8u(C_{mi^*}, C_r, C_{pj^*}, \boldsymbol{\psi}, \boldsymbol{\delta}) \quad (22)$$

$$l_2(\mathbf{x}) = u(C_{mi^*-1}, C_{pj^*}, \boldsymbol{\psi}, \boldsymbol{\delta}) - 0.8u(C_{mi^*}, C_r, C_{pj^*}, \boldsymbol{\psi}, \boldsymbol{\delta}) \quad (23)$$

$$l_3(\mathbf{x}) = u(C_{mi^*}, C_{pj^*+1}, \boldsymbol{\psi}, \boldsymbol{\delta}) - 0.8u(C_{mi^*}, C_r, C_{pj^*}, \boldsymbol{\psi}, \boldsymbol{\delta}) \quad (24)$$

$$l_4(\mathbf{x}) = u(C_{mi^*}, C_{pj^*-1}, \boldsymbol{\psi}, \boldsymbol{\delta}) - 0.8u(C_{mi^*}, C_r, C_{pj^*}, \boldsymbol{\psi}, \boldsymbol{\delta}) \quad (25)$$

where C_{mi^*} and C_{pj^*} are the set of coefficients that maximize the margins response for each of the C_r values.

A more convenient unconstrained formulation can be defined by adding a penalty term, as

$$\begin{aligned} U(\mathbf{x}) = -w_1[e_w(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta})][e_h(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta})] + w_2[e_{wa}(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta})] \\ + w_3[e_{ha}(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta})] + \gamma_0^l |L(\mathbf{x})|^2 \end{aligned} \quad (26)$$

where $\gamma_0^l \in \mathcal{R}$ is the penalty coefficient and $L(\mathbf{x})$ is a corner limits penalty function, defined as

$$L(\mathbf{x}) = \max\{0, l_1(\mathbf{x}), l_2(\mathbf{x}), l_3(\mathbf{x}), l_4(\mathbf{x})\} \quad (27)$$

Then, we aim at finding the optimal set of coefficients values \mathbf{x}^* by solving (17) with (26) as objective.

The combination of pattern search and the Nelder-Mead is a good approach to deal with our objective function (26) that contains many local minima. We start the optimization with pattern search, which serves for exploring the design space until finding a potential region where the global minimum is located. Then, the solution found by pattern search is used as seed for the Nelder-Mead method, which further minimizes the objective function for a more precise solution.

D. System Test Setup

The system under test is an Intel post-silicon validation platform. The PCIe link is exercised at the packet level with a protocol add-in test card which emulates the external device, as shown in Fig. 10. Measurements are based on the SMV

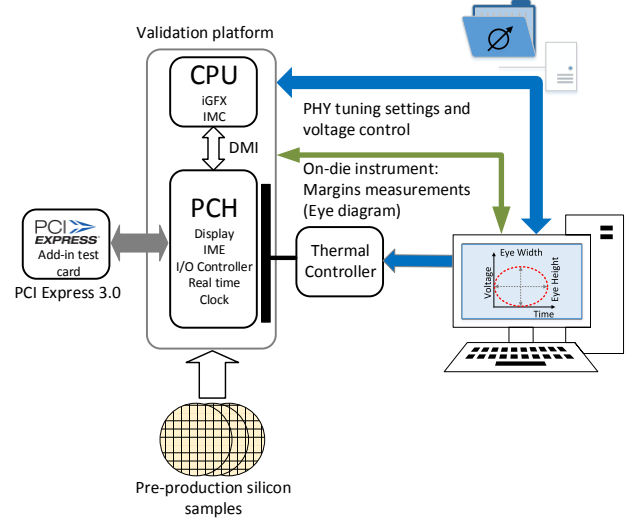


Fig. 10. PCI Express setup: an Intel server post-Si validation [22].

process. The optimization algorithm described in the previous section is implemented in Python, using the SciPy [24] modules for Nelder-Mead and pattern search algorithms.

E. Results

Through the optimization process defined in Section V.C, we arrive to a set of Tx and Rx coefficients in just 47 iterations, as shown in Fig. 11, which are executed in 4 hours.

A comparison on eye diagrams between the proposed methodology against the initial design and the exhaustive method is shown in Fig. 12. The optimized equalization coefficients yield an eye diagram with an e_h and e_w being now 30 ticks and 27 ticks, respectively, which corresponds to an improvement of 35% on eye diagram area as compared to that one with the initial coefficients. Even though the optimized coefficients show an eye diagram area decrease of 6% as compared to the exhaustive method, the efficiency of this approach is demonstrated by the reduction of the eye diagram asymmetries, and a significant time reduction in post-silicon validation. While the exhaustive method requires a few days for EQ maps data collection and analysis for a complete optimization (prone to human errors), the method proposed here can be completed in just 4 hours.

VI. HSIO RECEIVER COARSE SURROGATES MODELING

In [25], we propose a metamodeling approach, based on artificial neural networks (ANN), to efficiently simulate the silicon Rx equalizer. The model is generated using a frugal set of training data exploiting several DoE approaches to reduce the number of test cases. We evaluate the neural model performance by comparing with actual measured responses. The proposed methodology is illustrated by modeling two industrial HSIO topologies: USB3 Gen 1 and SATA Gen 3.

A. ANN Topology

Multilayer perceptrons are feedforward networks widely used as the preferred ANN topology. Since a 3-layer

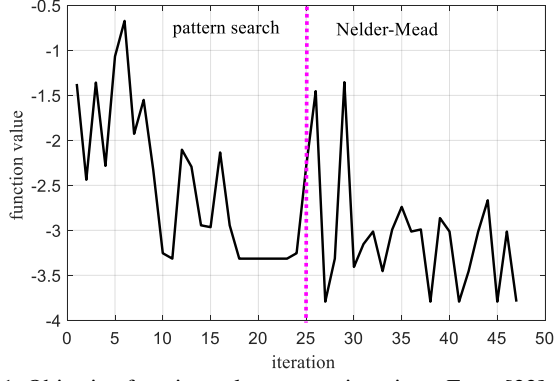


Fig. 11. Objective function values across iterations. From [22].

perceptron (3LP) is in principle sufficient for universal approximation [26], we use a 3LP to implement our neuromodel, with n inputs (equal to the number of Rx knobs), h hidden neurons, and m outputs (number of system responses of interest). The required complexity of the ANN, determined by h , depends on the generalization performance for a given set of training and testing data [27]. Following [28], we gradually increase h during training for regularization.

B. ANN Modeling and Training

Here we follow our work in [2] to define the corresponding objective function. The ANN is trained to find an optimal vector of weighting factors \mathbf{w} , such that the ANN response, denoted as \mathbf{R}_s , is as close as possible to the fine model response for all \mathbf{x} , $\boldsymbol{\psi}$, $\boldsymbol{\delta}$ in the region of interest,

$$\mathbf{R}_s(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta}, \mathbf{w}) \approx \mathbf{R}_f(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta}) \quad (28)$$

The ANN main input-output relationship is denoted as

$$\mathbf{R}_s = \mathbf{f}(\mathbf{x}) \quad (29)$$

We aim to develop a fast and accurate ANN model for \mathbf{f} by training the ANN with a set of measured learning data. The learning data are pairs of $(\mathbf{x}_L, \mathbf{t}_L)$, with $L = 1, 2, \dots, l$, where \mathbf{t}_L contains the desired outputs or targets (obtained from measurements) for the ANN model at the \mathbf{x}_L inputs, with l as the total number of learning samples. During training, we keep fixed the system at voltage/temperature (VT) nominal conditions and without changing the external devices. Under these conditions, $\boldsymbol{\psi}$ and $\boldsymbol{\delta}$ remain constant. Therefore, the ANN model during training is treated as

$$\mathbf{R}_{sL} = \mathbf{R}_s(\mathbf{x}_L, \mathbf{w}) \quad (30)$$

The ANN performance during training is evaluated by computing the difference between ANN outputs and the targets for all the learning samples,

$$\mathbf{E}_L(\mathbf{w}) = \mathbf{R}_{sL}(\mathbf{x}_L, \mathbf{w}) - \mathbf{t}_L \quad (31)$$

where \mathbf{E}_L is the learning error matrix.

Following [28], the problem of training the ANN is formulated as

$$\mathbf{w} = \arg \min_{\mathbf{w}} \|\mathbf{E}_L(\mathbf{w})\|_F \quad (32)$$

To control the generalization performance while solving (32), we use T testing base points (\mathbf{x}_T) not used during training. The scalar learning and testing errors are given by

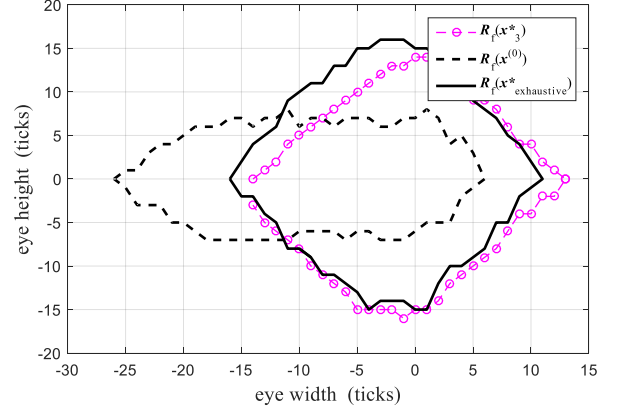


Fig. 12. Eye diagram results: comparing the proposed methodology ($\mathbf{R}_f(\mathbf{x}^*)$) against the initial design ($\mathbf{R}_f(\mathbf{x}^{(0)})$) and the exhaustive method ($\mathbf{R}_f(\mathbf{x}^*_{\text{exhaustive}})$). From [22].

$$\varepsilon_L = \|\mathbf{R}_{sL}(\mathbf{x}_L, \mathbf{w}) - \mathbf{R}_{tL}\|_F \quad (33)$$

$$\varepsilon_T = \|\mathbf{R}_{sT}(\mathbf{x}_T, \mathbf{w}) - \mathbf{R}_{tT}\|_F \quad (34)$$

where \mathbf{R}_{tT} and \mathbf{R}_{sT} are the output matrices of the fine model and ANN model, respectively, at the T testing base points, and \mathbf{R}_{tL} is the fine model response at the L learning base points.

The 3LP is trained by using the Bayesian regularization [29] method available in MATLAB Neural Network Toolbox. The algorithm for training the ANN is shown in [25]. We first define the learning ratio to split the pairs of inputs and targets into the learning and testing datasets. Then, we use a decoupling network process with initial set of inputs and outputs to compute initial weighting factors \mathbf{w}_0 and corresponding initial error $\varepsilon_T^{\text{old}}$. We start training the 3LP with just one hidden neuron ($h = 1$), and calculate the corresponding learning and testing errors. We keep increasing the complexity of the ANN (h) until the generalization performance starts to deteriorate [25].

C. Experimental System Configuration and DoE Approaches

The system under test is a server post-silicon validation platform, comprised of a CPU and a PCH. Within the PCH, our methodology was tested on two different HSIO links: USB3 Super-speed Gen 1 and SATA Gen 3. The measurement system is based on the SMV process.

We employ three different DoE techniques to explore the desired solution space with a reduced number of test cases. For each test case, we use seven input variables that represent Rx knobs ($n = 7$), which are settings used in three main Rx circuitry blocks (CTLE, VGA, and CDR), and then we retrieve the eye measurements from the system under test. The employed DoE techniques are: 1) Box Behnken (BB), which is type of second order response surface methodology (RSM) [30], using 62 experiments; 2) orthogonal arrays (OA) [31], using an $L_{27}(3^9)$ array in order to capture non-linear effects in the objective function by only running 27 experiments; and 3) Sobol [32] low-discrepancy sequence to sample the solution space. Given the quasi-Monte Carlo sampling approach of

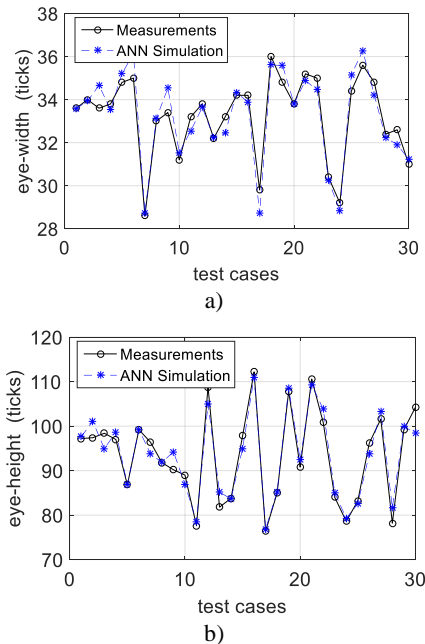


Fig. 13. Neural model generalization performance using Sobol50 for a) SATA eye width; b) SATA eye height.

Sobol, the solution space is better explored as the number of samples increases, at the expense of increasing test time on the real system. Therefore, we use three different Sobol DoE, denoted as Sobol50, Sobol100, and Sobol150, with 50, 100, and 150 samples, respectively.

System margin testing is very time consuming when running many test cases for PHY tuning. A single test case with 3 repetitions can take up 20 minutes, and then running a Sobol150 can take up 50 hours of testing for a single VT corner. The objective of comparing several DoEs is to find a suitable sampling strategy that provides adequate ANN model performance with the least amount of testing time.

D. Neural Modeling Results

Comparing the generalization error of the already trained neural model (at \mathbf{w}^*) for different DoEs, we found that the three Sobol cases provide the best generalization performance, and the best accuracy is achieved with Sobol150 (as expected). However, Sobol50 is able to achieve acceptable accuracy with only 50 samples. Considering the learning performance of the neural training algorithm for SATA. The best performance is achieved with $h = 3$ for the eye width ANN, achieving a maximum relative learning error of 3.65%, and 7.63% for the relative testing error. For the eye height ANN, best performance is achieved with $h = 4$, yielding 7.98% of learning error and 6.75% of testing error. Thus, the metamodels are able to reach above 90% of accuracy with these limited sampling points.

The neural model response at \mathbf{w}^* and $h = 3$ for e_w and $h = 4$ for e_h from Sobol50 is compared in Fig. 13a and Fig. 13b, respectively, with the fine model (real measurements), by using 30 testing base points not used during training, in order to test the generalization performance. It is observed that the

neural model effectively simulates the actual physical measurements with a total relative error of 1.7% for the e_w response and 2.5% for the e_h response. In other words, the ANN metamodel is able to predict margins with up to 95% of accuracy when using equalization settings not used during the ANN training. The technique can easily be applied to other interfaces, such as USB3 Super-speed Gen 1, as shown in [25].

VII. HSIO RECEIVER EQUALIZATION BY SPACE MAPPING OPTIMIZATION

In [33], we reported how the Broyden-based input space mapping (SM) algorithm, better known as aggressive SM (ASM) [34], [35], is used for the first time in HSIO PHY tuning optimization. Our SM approach takes advantage of a coarse surrogate model developed following [36]. In our case, the fine model is a measurement-based post-silicon validation industrial platform, while the coarse model is based on a Kriging surrogate technique. Our approach is illustrated by optimizing the PHY tuning Rx equalizer settings for a SATA Gen 3 channel topology.

A. Broyden-based Input Space Mapping

SM optimization methods belong to the general class of surrogate-based optimization algorithms [37]. They are specialized on the efficient optimization of computationally expensive models. The most widely used SM approach to efficient design optimization is the ASM or Broyden-based input space mapping algorithm [35]. ASM efficiently finds an approximation of the optimal design of a computationally expensive model (fine model) by exploiting a fast but inaccurate surrogate representation (coarse model) [35]. ASM aims at finding a solution that makes the fine-model response close enough to the desired response.

B. Fine Model

Our fine model is an Intel server post-silicon validation platform, as shown in Fig. 14. Within the PCH, our methodology is applied to a HSIO link SATA Gen3 [38]. The measurement system is based on the SMV process. We follow our work in [2] to define the corresponding objective function.

We use five input variables that represent the SATA Rx PHY tuning coefficients, which are settings used in three main Rx circuitry blocks (CTLE, VGA, and CDR). $e_w \in \mathfrak{R}$ and $e_h \in \mathfrak{R}$ are obtained from measured parameters,

$$e_w(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta}) = e_{wr}(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta}) + e_{wl}(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta}) \quad (35)$$

$$e_h(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta}) = e_{hh}(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta}) + e_{hl}(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta}) \quad (36)$$

where $e_{wr} \in \mathfrak{R}$ and $e_{wl} \in \mathfrak{R}$ are the eye width-right and eye width-left measured parameters, respectively, and $e_{hh} \in \mathfrak{R}$ and $e_{hl} \in \mathfrak{R}$ are the eye height-high and eye height-low parameters, respectively.

C. Coarse Model

Here, we follow our work in [36] to develop a coarse

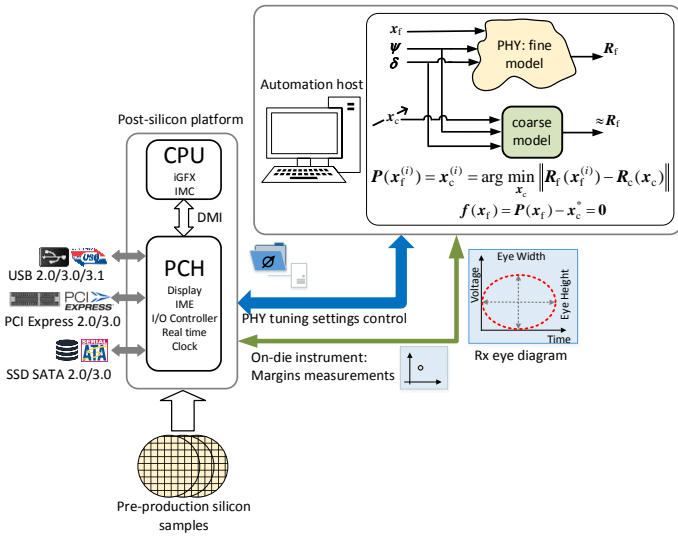


Fig. 14. An Intel server post-Si validation platform exploiting Broyden-based input space mapping design optimization. From [33].

surrogate model for a HSIO link SATA Gen3. By using the PHY tuning setting coefficients as inputs \mathbf{x} and the corresponding eye height and width as outputs \mathbf{R}_c , we select a Kriging surrogate modeling technique [12] with a Sobol [32] DoE approach with only 50 samples.

D. Objective Function

We want to find the optimal set of PHY tuning settings \mathbf{x} that maximize the functional eye diagram area. Therefore, our objective function is given by

$$u(\mathbf{x}) = -[e_w(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta})][e_h(\mathbf{x}, \boldsymbol{\psi}, \boldsymbol{\delta})] \quad (37)$$

During SM optimization, both $\boldsymbol{\psi}$ and $\boldsymbol{\delta}$ are kept fixed.

E. ASM Optimization Results

After applying the Broyden-based input SM algorithm [35], we arrive to a space-mapped solution, \mathbf{x}^{SM} , in just 6 iterations (or fine model evaluations). The set of Rx EQ coefficients contained in \mathbf{x}^{SM} makes the measured SATA Rx inner eye height and width of the PCH as open as that one predicted by the optimized coarse surrogate model. The SM solution (\mathbf{x}^{SM}) found makes an improvement of 85% on the fine model eye diagram area as compared to that one with the initial settings ($\mathbf{x}_c^{(0)}$), and a 33% improvement as compared to that one with the optimal coarse model solution (\mathbf{x}_c^*), as shown in Fig. 15.

The efficiency of this approach is also demonstrated by a very significant time reduction in post-silicon validation and PHY tuning Rx equalization. While the traditional industrial process requires days for a complete empirical optimization, the method proposed here can be completed in a few hours. The technique can easily be applied to other interfaces like USB and PCI express.

VIII. DISCUSSION AND CONCLUSION

In this paper, we proposed direct and surrogate-based

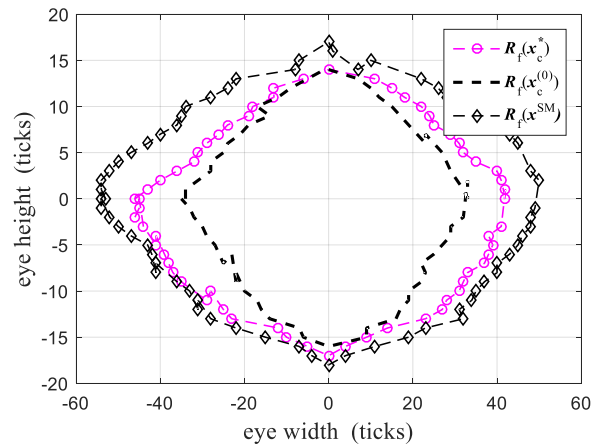


Fig. 15. Comparison between the system fine model responses at the initial Rx EQ coefficients, $\mathbf{x}_c^{(0)}$, at the optimal coarse model solution, \mathbf{x}_c^* , and at the space-mapped solution found, \mathbf{x}_c^{SM} . From [33].

optimization methods, including space mapping, based on suitable objective functions to efficiently tune the Tx and Rx equalizers coefficients. The experimental results, based on real industrial validation platforms, demonstrated the efficiency of the proposed methods, showing a substantial improvement as compared with the current industrial practice, and accelerating the typical required time for equalizers tuning.

A holistic optimization approach that merges system margining and jitter tolerance measurements for PHY tuning was demonstrated. The experimental results demonstrated the efficiency of the method to deliver optimal margins while ensuring jitter tolerance compliance, showing a substantial improvement for both system margins and jitter tolerance as compared with the current industrial practice, and dramatically accelerating the typical time required for PHY tuning.

It was also proposed a direct optimization approach based on a suitable objective function formulation to efficiently tune the Tx FIR filter for the Ethernet SFP+ interface. The optimized coefficients were evaluated by measuring the real eye diagram of the physical system, showing a great mitigation of the ISI effects, and accelerating the typical required time for Tx coefficients tuning.

A direct optimization approach for PCIe link equalization was also proposed based on a suitable objective function formulation to efficiently tune the Tx FIR filter and Rx CTLE EQ coefficients to mitigate ISI and other undesired channel effects, and successfully comply with the PCIe specification. The optimized EQ coefficients were evaluated by measuring the real eye diagram of the physical system, demonstrating a great mitigation of the ISI and channel effects, and significantly enhancing current PCIe Tx/Rx tuning industrial practices in post-silicon validation.

A metamodeling technique based on artificial neural networks was also presented to efficiently simulate the effects of the Rx EQ circuitry in industrial HSIO links. Through the proposed neural modeling procedure, an efficient surrogate model is found that approximates the system with a reduced set of testing and training data.

Finally, it was also described how the Broyden-based input space mapping algorithm can be used to efficiently optimize the PHY tuning Rx equalizer settings by exploiting a low-cost low-precision Kriging model, and a measurement-based post-silicon validation platform as the fine model. The experimental results, based on a real industrial validation platform, demonstrated the efficiency of the method, showing a substantial performance improvement and a dramatic acceleration of the typical required time for PHY tuning.

Ultimately, the present paper is based on the doctoral dissertation [39].

REFERENCES

- [1] J. Keshava, N. Hakim, and C. Prudvi, "Post-silicon validation challenges: How EDA and academia can help," in *Proc. Design Automation Conf. (DAC)*, Anaheim, CA, June 2010, pp. 3–7.
- [2] F. E. Rangel-Patiño, A. Viveros-Wacher, J. E. Rayas-Sánchez, I. Durón-Rosales, E. A. Vega-Ochoa, N. Hakim and E. López-Miralrio, "A holistic formulation for system margining and jitter tolerance optimization in industrial post-silicon validation," *IEEE Trans. Emerging Topics Computing*, vol. 6, 2018. (published online: 29 Sep. 2017).
- [3] W. Yao, Y. Shi, L. He, S. Pamarti, "Joint design-time and post-silicon optimization for digitally tuned analog circuits," in *2009 IEEE/ACM Int. Conf. on CAD*, San Jose, CA, Nov. 2009, pp. 725-730.
- [4] H. Huang and E. K. F. Lee, "Design of low-voltage CMOS continuous time filter with on-chip automatic tuning," *IEEE J. Solid-State Circuits*, vol. 36, no. 8, pp. 1168-1177, Aug. 2001.
- [5] G. Miller, M. Timko, H.-S. Lee, E. Nestler, M. Mueck, and P. Ferguson, "Design and modeling of a 16-bit 1.5msps successive approximation adc with non-binary capacitor array," in *Proc. Int. Great Lakes Symp. on VLSI*, Washington, D. C. April 2003, pp. 161 - 164
- [6] F. Rangel-Patino, A. Viveros-Wacher, J. E. Rayas-Sánchez, E. A. Vega-Ochoa, I. Duron-Rosales, and N. Hakim, "A holistic methodology for system margining and jitter tolerance optimization in post-silicon validation," in *IEEE MTT-S Latin America Microw. Conf. (LAMC-2016)*, Puerto Vallarta, Mexico, Dec. 2016, pp. 1-4.
- [7] C. Gu, "Challenges in post-silicon validation of high-speed I/O links," in *Proc. Intl. Conf. on Computer-Aided Design (ICCAD), IEEE/ACM*, San Jose, CA, Nov. 2012, pp. 547-550.
- [8] A. Viveros-Wacher et al, "SMV methodology enhancements for high speed IO links of SoCs", in *IEEE VLSI Test Symposium (VTS)*, Napa, CA, Apr. 2014, pp. 1-5.
- [9] Y. Fan, Z. Zilic, "Accelerating jitter tolerance qualification for high speed serial interfaces", in *IEEE Int. Symp. on Quality Electronic Design*, San Jose, CA, Mar. 2009, pp. 360-365.
- [10] USB Org. (2016). *Universal Serial Bus Revision 3.1 Specification* [Online]. Available: <http://www.usb.org/developers/doc>
- [11] F. E. Rangel-Patiño, J. L. Chávez-Hurtado, A. Viveros-Wacher, J. E. Rayas-Sánchez, and N. Hakim, "System margining surrogate-based optimization in post-silicon validation," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 9, pp. 3109-3115, Sep. 2017..
- [12] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments," *Statistical Science*, vol. 4, no. 4, pp. 409-423, Nov. 1989.
- [13] S. N. Lophaven, H. B. Nielsen, J. Søndergaard, "Aspects of the Matlab toolbox DACE," Internal Report *IMM-TR-2002-13*, Technical University of Denmark, Lyngby, 2002.
- [14] ANRITSU (2010). *Understanding Eye Pattern Measurements*, application note no. 11410-00533 (revision A) [Online]. Available: <https://www.anritsu.com>
- [15] *IEEE Standard for Ethernet*. IEEE Standard 802.3-2015.
- [16] K. S. Reddy, S. K. Sahoo, "An approach for FIR filter coefficient optimization using differential evolution," in *Int. J. Electron. Commun. (AEÜ)*, 2015, vol. 65, no. 1, pp. 101–108.
- [17] M. Kumm, K. Möller, P. Zipf, "Reconfigurable FIR Filter Using Distributed Arithmetic on FPGAs," in *IEEE Int. Symp. Circuits and Systems (ISCAS)*, Kuala Lumpur, Malaysia, May 2013.
- [18] Y. S. Cheng, Y. C. Lai, and R. B. Wu, "Optimization of FIR filter to improve eye diagram for general transmission line systems," in *Design, Automation & Test in Europe Conf. & Exhibition*, Dresden, Germany, Apr. 2010, pp. 1321-1324.
- [19] I. Duron-Rosales, F. Rangel-Patino, J. E. Rayas-Sánchez, J. L. Chávez-Hurtado, and N. Hakim, "Reconfigurable FIR filter coefficient optimization in post-silicon validation to improve eye diagram for optical interconnects," in *Int. Caribbean Conf. Devices, Circuits, and Systems (ICDCS-2017)*, Cozumel, Mexico, Jun. 2017, pp. 85-88.
- [20] A. Wilen, J. Shade, and R. Thornburg, *Introduction to PCI Express: A Hardw. and Softw. Developer's Guide*. Hillsboro, OR: Intel Press, 2003.
- [21] PCI SIG Org. (2017), *PCI Express® Base Specification Revision 4.0 Version 1.0* [Online]. Available: <https://pcisig.com/specifications>.
- [22] F. E. Rangel-Patiño, J. E. Rayas-Sánchez, E. A. Vega-Ochoa, and N. Hakim, "Direct optimization of a PCI Express link equalization in industrial post-silicon validation," in *IEEE Latin American Test Symp. (LATS 2018)*, Sao Paulo, Brazil, Mar. 2018, pp. 1-6.
- [23] D. R. Stauffer, J. T. Mechler, M. Sorna, K. Dramstad, C. R. Ogilvie, A. Mohammad, and J. Rockrohr, *High Speed SERDES Devices and Applications*, New York, NY: Springer, 2008.
- [24] SciPy.org, PCI SIG Org. (2017), *Open-source Softw. for Math., Science, Eng.* [Online]. Available: <https://docs.scipy.org/doc/scipy/reference>.
- [25] F. E. Rangel-Patiño, J. E. Rayas-Sánchez, A. Viveros-Wacher, J. L. Chávez-Hurtado, E. A. Vega-Ochoa, and N. Hakim, "Post-silicon receiver equalization metamodeling by artificial neural networks," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 37, 2018. (published online: 8 May 2018).
- [26] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359-366, May 1989.
- [27] J. E. Rayas-Sánchez, *Neural Space Mapping Methods for Modeling and Design of Microw. Circuits*, Ph.D. Thesis, Dept. of Electrical and Comp. Eng., McMaster University, Hamilton, Canada, 2001.
- [28] J. E. Rayas-Sánchez and V. Gutiérrez-Ayala, "EM-based Monte Carlo analysis and yield prediction of microwave circuits using linear-input neural-output space mapping," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 12, pp. 4528-4537, Dec. 2006.
- [29] D. J. C. MacKay, "Bayesian interpolation," *Neural Computation*, vol. 4, no. 3, pp. 415–447, 1992.
- [30] C. F. J. Wu and M. Hamada, *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York, NY: Wiley, 2000.
- [31] Y. P. Chang et al, "Design of discrete-value passive harmonic filters using sequential neural-network approximation and orthogonal array", in *IEEE PES Transm. Distrib. Conf. & Expo.: Asia Pacific*, Dalian, China, Aug. 2005, pp. 1-6.
- [32] I. M. Sobol, "On the distribution of points in a cube and the approximate evaluation of integrals," *U.S.S.R. Computational Mathematics and Mathematical Physics*, vol. 7, no. 4, pp. 86-112, 1967.
- [33] F. E. Rangel-Patiño, J. E. Rayas-Sánchez, A. Viveros-Wacher, E. A. Vega-Ochoa, and N. Hakim, "High-speed links receiver optimization in post-silicon validation exploiting Broyden-based input space mapping," in *IEEE MTT-S Int. Conf. Num. EM Mutiphys. Model. Opt. RF, Microw., Terahertz App. (NEMO-2018)*, Reykjavik, Iceland, Aug. 2018, pp. 1-3.
- [34] J. W. Bandler, R. M. Biernacki, S. H. Chen, R. H. Hemmers, and K. Madsen, "Electromagnetic optimization exploiting aggressive space mapping," *IEEE Trans. Microw. Theory Techn.*, vol. 41, no. 12, pp. 2874-2882, Dec. 1995.
- [35] J. E. Rayas-Sánchez, "Power in simplicity with ASM: tracing the aggressive space mapping algorithm over two decades of development and engineering applications," *IEEE Microw. Magazine*, vol. 17, no. 4, pp. 64-76, Apr. 2016.
- [36] F. Rangel-Patino, J. L. Chávez-Hurtado, A. Viveros-Wacher, J. E. Rayas-Sánchez, and N. Hakim, "Eye diagram system margining surrogate-based optimization in a server silicon validation platform," in *Europ. Microw. Conf. (EuMC-2017)*, Nuremberg, Germany, Oct. 2017, pp. 540-543.
- [37] A. J. Booker, J. E. Dennis Jr., P. D. Frank, D. B. Serafini, V. Torczon, and M. W. Trosset, "A rigorous framework for optimization of expensive functions by surrogates," *Struct. Optim.*, vol. 17, pp. 1–13, Feb. 1999
- [38] SATA Org. (2016), *Serial Advanced Technology Attachment 3.2 Specification* [Online]. Available: <http://www.sata-io.org/>
- [39] F. E. Rangel-Patiño, *Transmitter and Receiver Equalizers Optimization Methodologies for High-Speed Links in Industrial Computer Platforms Post-Silicon Validation*, PhD Thesis, Dept. of Electronics, Systems, and Informatics, ITESO, Tlaquepaque, Mexico, 2018.