PCIe Gen5 Physical Layer Equalization Tuning by Using K-means Clustering and Gaussian Process Regression Modeling in Industrial Post-silicon Validation

Francisco E. Rangel-Patiño[#*1], Andres Viveros-Wacher[#2], Chintan Rajyaguru[$3], Edgar A. Vega-Ochoa[#4], Sofia D. Rodriguez-Saenz[#5], Johana L. Silva-Cortes[#6], Hemanth Shival[$7], and José E. Rayas-Sánchez[*8],
[#] Intel Corp. Zapopan, Jalisco, 45019 Mexico
[$] Intel Corp. Folsom, CA, 95630 USA
[*] Department of Electronics, Systems, and Informatics, ITESO – The Jesuit University of Guadalajara, Tlaquepaque, Jalisco, 45604 Mexico
[1]francisco.rangel, [2]andres.viveros.wacher, [3]chintan.rajyaguru, [4]edgar.vega.ochoa, [5]sofia.d.rodriguez.saenz, [6]johana.l.silva.cortes, [7]hemanth.shival{@intel.com}, [8]erayas@iteso.mx

***Abstract*—Peripheral component interconnect express (PCIe) is a high-performance interconnect architecture widely adopted in the computer industry. The continuously increasing bandwidth demand from new applications has led to the development of the PCIe Gen5, reaching data rates of 32 GT/s. To mitigate undesired channel effects due to such high-speed, the PCIe specification defines an equalization process at the transmitter (Tx) and the receiver (Rx). Current post-silicon validation practices consist of finding an optimal subset of Tx and Rx coefficients by measuring the eye diagrams across different channels. However, these experiments are very time consuming since they require massive lab measurements. In this paper, we use a K-means approach to cluster all available post-silicon data from different channels and feed those clusters to a Gaussian process regression (GPR)-based metamodel for each channel. We then perform a surrogate-based optimization to obtain the optimal tuning settings for the specific channels. Our methodology is validated by measurements of the functional eye diagram of an industrial computer platform.**

***Keywords*—clustering, equalization, equalization maps, eye-diagram, FIR, GPR, HSIO, high-speed links, metamodels, optimization, PCIe, post-silicon validation, receiver, signal integrity, transmitter, tuning**

I. Introduction

Post-silicon validation of modern embedded digital systems has become a huge task due to the increasing system complexity. It requires tens or hundreds of person-hours and needs the computing power of hundreds of workstations [1]. A significant portion of the circuits to be validated in modern microprocessors corresponds to high-speed input/output (HSIO) links, for which direct and surrogate-based optimization methods, including space mapping, have been employed to efficiently tune the transmitter (Tx) and receiver (Rx) equalizers [2].

Peripheral component interconnect express (PCIe) is one of the most complex HSIO interfaces and the primary interface for a host central processing unit (CPU) to connect with input/output (I/O) devices. PCIe has been continuously evolving and the new PCIe Gen6 specification, released in 2021, has reached a data rate of 64 giga-transfers per second (GT/s). However, as transmission speeds increase, undesired channel effects such as reflections, crosstalk, jitter, and inter-symbol interference (ISI), are more severe, causing the signals to become more susceptible to errors [3]. Additionally, PCIe channels are bandwidth-limited by default, causing large signal attenuation at high frequencies. This generates distortion and spreading of the transmitted signal over multiple symbols, exacerbating ISI, which can make the signal unreadable at the Rx, producing bit errors. The most practical solution to this problem is signal conditioning to open the eye diagram [4].

PCIe specification defines an adaptive mechanism for equalization (EQ) to determine the optimum values of the Tx and Rx EQ coefficients within a fixed time limit, across the allowed channel types. The most widely used current method consists of using maps of EQ coefficients, which are obtained from massive eye diagram measurements. The EQ maps are used to characterize the PCIe link across different channel losses and devices. Once the full characterization is completed, the best Tx EQ values are selected based on the input of an experienced validation engineer. This is a very time-consuming process and prone to human errors [5]. Direct measurement-based numerical optimization approaches have also been employed for PCIe EQ tuning [4],[6].

Machine learning algorithms are useful to build statistical models from examples, which are then used to make predictions when faced with cases not seen before [7]. Unsupervised machine learning algorithms are designed to learn patterns from untagged data. On the other hand, supervised machine learning models are trained to predict outputs from a given set of inputs. The large volume of data generated from typical post-silicon testing suggests the application of machine learning techniques [8] to identify underlying patterns, such as channel effects on the analog behaviour of the HSIO link.

In this paper, we first use unsupervised machine learning techniques to cluster all available post-silicon data from different channels, dividing them into distinct sets of channel conditions. We then develop statistical supervised machine learning models, based on Gaussian process regression (GPR), to predict the eye diagram margins within each data subset. We finally optimize the GPR-based models to obtain the optimal

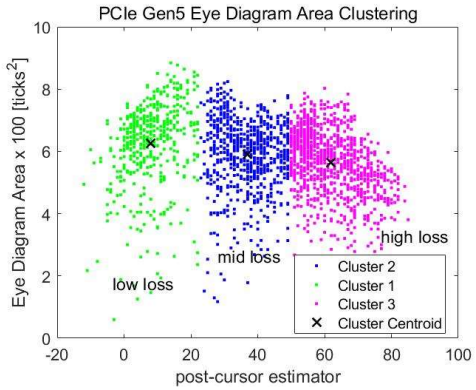


Fig. 1. K-means clustering forms 3 clusters of eye diagram margins area values, divided by low-, mid-, and hi-loss channels. Each colored dataset is settled into one cluster segment. Cluster centroids are also indicated.

tuning settings for the specific channels. Our proposed method is validated by measurements of the functional eye diagram of an actual industrial computer platform.

II. K-means Clustering

Clustering is an unsupervised machine learning approach to find subsets, or clusters, of data based on the similarity between data points in the same cluster, and dissimilarity with data points in other clusters [9]. Clustering methods are divided based on how similarity and dissimilarity are measured, and as such, can be classified into hierarchical and partitional methods. Partitional methods suppose that the dataset can be divided into finite clusters by measuring distance.

K-means is one of the most widely used partitional clustering methods. With a pre-defined k number of cluster centers (or centroids), it iteratively establishes the centroid coordinates and measure the Euclidean distance between each data point and each centroid, to group the data points based on minimum distance [10],[11]. In our work, we employ k-means to cluster margin data points based on channel loss similarities.

III. Gaussian Process Regression Modeling

GPR is a non-parametric Bayesian approach to regression. Whereas many popular supervised machine learning algorithms aim to learn exact outputs for a given set of inputs, GPR looks to infer a probability distribution over all possible output values of the model. Additionally, GPR is non-parametric since rather than calculating a distribution of parameters for a specific function, it calculates the distribution over admissible functions that fit the data [12]. This is of extreme value for modelling non-deterministic data, *e.g.*, obtained from physical measurements subject to statistical uncertainty and varying operating or environmental conditions [13], as in the case of typical post-silicon measurements of eye diagram margins. In contrast to the work in [14], where several deterministic surrogate models are developed from eye diagram margins obtained for a SATA Gen 3 HSIO interface, here we use GPR to create non-parametric, probabilistic models based on PCIe eye diagram margins data.

IV. Objective Function Formulation and Optimization

We aim at finding the optimal set of EQ coefficients $\boldsymbol{x}^*$ that

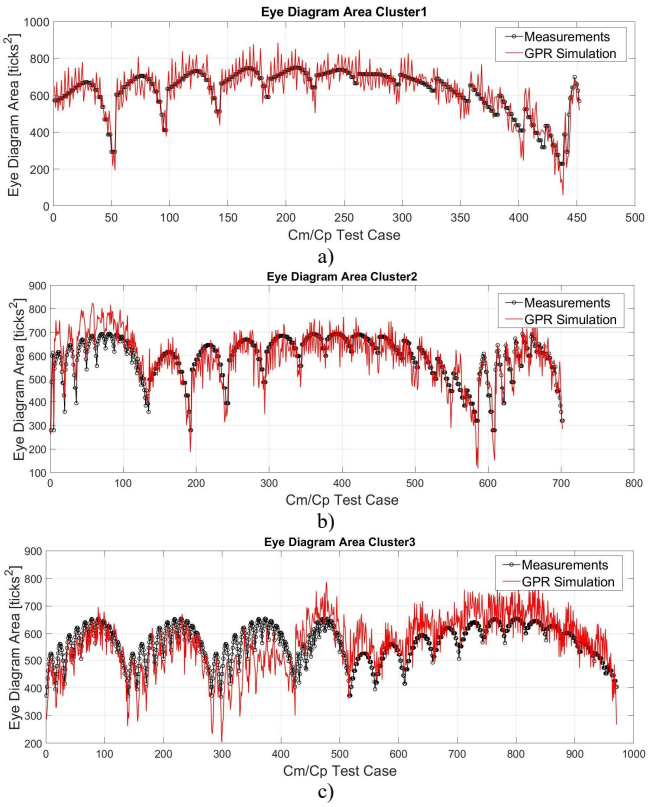


Fig. 2. Comparing the resultant GPR models responses versus actual measurements across all possible channels and C_m/C_p combinations: a) cluster #1 (low-loss), b) cluster #2 (mid-loss), c) cluster #3 (hi-loss). See Fig. 1.

maximizes the eye diagram based on the margin response. We follow [4] to define an initial objective function as

$$u(\boldsymbol{x}) = -[e_w(\boldsymbol{x})][e_h(\boldsymbol{x})] \tag{1}$$

where e_w and e_h are the width and height, respectively, of the eye diagram margins. The eye width and height are functions of the Tx finite impulse response (FIR) filter pre-cursor (C_m) and post-cursor (C_p) EQ coefficient values (integer numbers), contained in vector $\boldsymbol{x}$.

Additionally, we need to ensure that the system margin response at $\boldsymbol{x}^*$ is not too sensitive, *i.e.*, $\boldsymbol{x}^*$ should lie in a sufficiently flat region of the EQ map space [4]. In order to satisfy this requirement, the four margin responses around $u(\boldsymbol{x}^*)$ must be at least 80% of the value of $u(\boldsymbol{x}^*)$.

The new optimization problem can be defined through a constrained formulation, such that the optimal set of coefficients maximizes the system response without violating the lower bound of $0.8u(\boldsymbol{x}^*)$ in the vicinity,

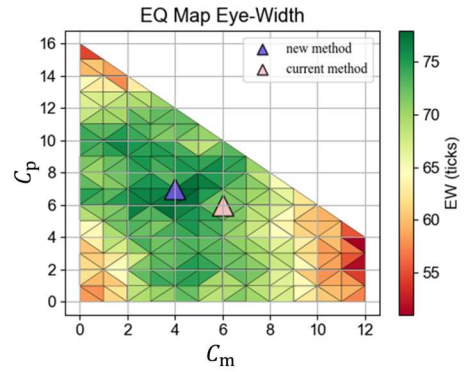
$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} u(\boldsymbol{x}) \tag{2}$$

$$\text{subject to } \; l_{11}(\boldsymbol{x}) \le 0,\, l_{12}(\boldsymbol{x}) \le 0,\, l_{21}(\boldsymbol{x}) \le 0,\, l_{22}(\boldsymbol{x}) \le 0$$

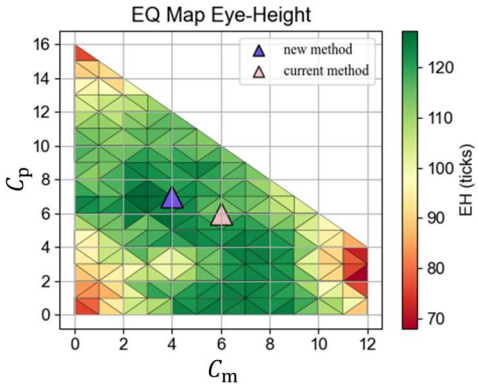
with matrix $\boldsymbol{I}(\boldsymbol{x})$ given by

$$\boldsymbol{I}(\boldsymbol{x}) = 0.8u(\boldsymbol{x})\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} - \begin{bmatrix}u(\boldsymbol{x}-[1\,0]^{\mathrm{T}}) & u(\boldsymbol{x}+[1\,0]^{\mathrm{T}})\\ u(\boldsymbol{x}-[0\,1]^{\mathrm{T}}) & u(\boldsymbol{x}+[0\,1]^{\mathrm{T}})\end{bmatrix} \tag{3}$$

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



a)



b)

Fig. 3. PCIe lane (16.8 dB) EQ map Tx FIR filter pre-cursor (C_m) and post-cursor (C_p) values. Comparing the solution found $\boldsymbol{x}^*$ by the proposed methodology (purple marker) against that one found by the current method (pink marker): a) eye width EQ map, b) eye height EQ map.

$$U(\boldsymbol{x}) = -[e_w(\boldsymbol{x})][e_h(\boldsymbol{x})] + L(\boldsymbol{x})\left[\frac{|u(\boldsymbol{x}^{(0)})|}{\max\{\boldsymbol{I}(\boldsymbol{x}^{(0)})\}}\right] \quad (4)$$

where $\boldsymbol{x}^{(0)}$ is the starting point and $L(\boldsymbol{x})$ is defined as

$$L(\boldsymbol{x}) = \max\{0, \max\{\boldsymbol{I}(\boldsymbol{x})\}\} \quad (5)$$

Our unconstrained objective reduces to

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} U(\boldsymbol{x}) \quad (6)$$

We aim at finding the optimal set of EQ coefficients values $\boldsymbol{x}^*$ by solving (6) using the gradient-free computationally efficient direct-search optimization method Nelder-Mead [15].

V. Results

We first cluster all available post-silicon data across different channels, dividing them into distinct sets of channel losses. We infer the amount of losses by using an intermediate parameter (post-cursor estimator) from the microcontroller algorithm inside the CPU, which follows the trend of the channel loss. After applying the K-means clustering algorithm, we are able to cluster all margin data points (area of the eye diagram margins) based on channel loss similarities, creating the three clusters shown in Fig. 1.

Then, the clustered data points are used to develop three GPR models, with the equalization coefficients as inputs and the area of the eye diagram margins as output. We evaluated the accuracy of the obtained GPR models by comparing with actual measured responses, using an average relative error defined as in [14]. The relative error for cluster 1 and 2 was 25%, and the error for cluster 3 was 30%. Fig. 2 shows the inherent large data variability, which causes these high average relative errors; however, it is also seen in Fig. 2 that the resultant GPR models are able to follow the overall trends of the actual measurements. While these models could not be regarded as highly accurate, they are good enough to perform effective and inexpensive surrogate-based optimization (SBO).

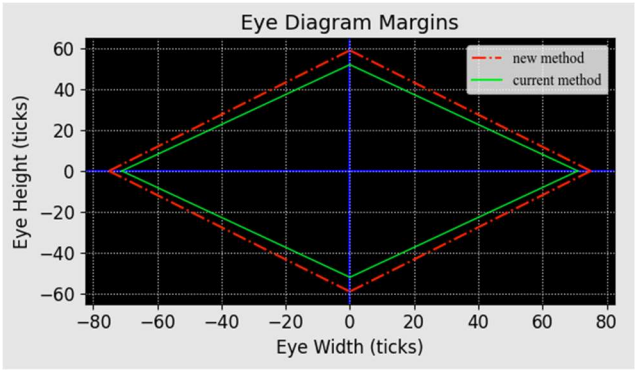


Fig. 4. PCIe lane (16.8 dB) eye diagram margins comparing the proposed methodology against the current method.

We next perform a SBO with these GPR models using the formulation in Section IV, to obtain the optimal tuning Tx equalizer settings. Fig. 3 shows the EQ maps of one of the mid-losses PCIe lanes (16.8 dB at 16 GHz) considering all Tx FIR filter C_m and C_p feasible combinations, comparing the position of the solution obtained using the current method against the position of $\boldsymbol{x}^*$ obtained from proposed methodology. It is evident that $\boldsymbol{x}^*$ obtained with the proposed methodology is in a more suitable region of the EQ map space and better satisfies the criteria of the four margin responses around $u(\boldsymbol{x}^*)$ to be at least 80% of the value of $u(\boldsymbol{x}^*)$.

Finally, we validate the SBO results by measuring the PCIe Gen5 link Rx inner eye height/width at $\boldsymbol{x}^*$ on the real validation platform with a PCIe Gen5 test card. The results, shown in Fig. 4, indicate an improvement of 20% on eye diagram area as compared to the tuning settings with the current method, demonstrating the effectiveness of our approach. Moreover, a significant time reduction in post-silicon validation is achieved with the proposed new methodology. While the current method requires days of effort on data collection and EQ maps analysis for a complete optimization (prone to human errors), the proposed new method can be completed in a few hours.

VI. Conclusion

We proposed an SBO approach for PCIe Gen5 link equalization based on a suitable objective function formulation to efficiently tune the Tx FIR filter coefficients to maximize the area of the eye diagram margins and successfully comply with the PCIe specification. We use machine learning techniques to cluster all available post-silicon data from different channels and feed those clusters to a GPR-based metamodel for each channel. The optimized EQ coefficients were validated by

measuring the eye diagram margins of the physical system, demonstrating a significant increase in eye diagram margins area and accelerating the typical long time for EQ tuning.

REFERENCES

[1] F. E. Rangel-Patino, 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. 8, no. 2, pp. 453-463, Apr.-Jun. 2020.

[2] F. E. Rangel-Patiño, J. E. Rayas-Sánchez, and N. Hakim, "Transmitter and receiver equalizers optimization methodologies for high-speed links in industrial computer platforms post-silicon validation," in *Int. Test Conf. (ITC-2018)*, Phoenix, AZ, Oct. 2018, pp. 1-10.

[3] M. Li, *Jitter, Noise, and Signal Integrity at High-Speed*. Boston, MA: Prentice Hall, 2007.

[4] 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.

[5] 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 Integ. Circ. Syst.*, vol. 38, no. 4, pp. 733-740, Apr. 2019.

[6] R. J. Ruiz-Urbina, F. E. Rangel-Patiño, J. E. Rayas-Sánchez, E. A. Vega-Ochoa, and O. Longoria-Gándara, "Transmitter and receiver equalizers optimization for PCI Express Gen6.0 based on PAM4," in *IEEE MTT-S Latin America Microw. Conf. (LAMC)*, Cali, Colombia, May 2021, pp. 1-4.

[7] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*. London, UK: Elsevier Academic Press. 2015.

[8] L. Wang, "Experience of data analytics in EDA and test—principles, promises, and challenges," *IEEE Trans. Computer-Aided Des. Integr. Circ. Syst.*, vol. 36, no. 6, pp. 885-898, Jun. 2017.

[9] K. P. Sinaga and M. -S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716-80727, May 2020.

[10] S. Kapil, M. Chawla, and M. D. Ansari, "On K-means data clustering algorithm with genetic algorithm," in *Int. Conf, Parallel, Distributed Grid Computing (PDGC)*, Waknaghat, India, Dec. 2016, pp. 202-206.

[11] B. Zhang, "Regression clustering," in *Third IEEE Int. Conf. Data Mining*, Melbourne, FL, Nov. 2003, pp. 451-458.

[12] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, Massachusetts: MIT Press, 2006.

[13] J. E. Rayas-Sánchez, S. Koziel, and J. W. Bandler, "Advanced RF and microwave design optimization: a journey and a vision of future trends," *IEEE J. of Microwaves*, vol. 1, no. 1, pp. 481-493, Jan. 2021.

[14] 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.

[15] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder-Mead simplex method in low dimensions," *SIAM J. Optimization*, vol. 9, no. 1, pp. 112-147, 1998.