This is the author's version of an article that has been published in this conference. Changes were made to this version by the publisher prior to publication. The final version is available at http://dx.doi.org/10.1109/LAMC59011.2023.10375599

# PCIe Gen6 Physical Layer Equalization Tuning by Using Unsupervised and Supervised Machine Learning Techniques

Francisco E. Rangel-Patiño[#*], Andres Viveros-Wacher[#], Sofia D. Rodriguez-Saenz[#], José E. Rayas-Sánchez[*], Edgar A. Vega-Ochoa[#], and Hemanth Shival[$]
[#] Intel Corp. Zapopan, Jalisco, 45019 Mexico, francisco.rangel@intel.com
[$] Intel Corp. Folsom, CA, 95630 USA
[*] ITESO – The Jesuit University of Guadalajara, Tlaquepaque, Jalisco, 45604 Mexico

***Abstract*—Ever faster applications triggered the development of the new PCIe Gen6 specification, reaching 64 GT/s data rates with PAM4 modulation. This brings new challenges for the physical channel design, where equalization (EQ) plays a key role. PCIe specification defines an EQ process at the transmitter (Tx) and the receiver (Rx). Current post-silicon validation practices consist of finding optimal subsets of Tx and Rx coefficients by measuring the eye diagram at the Rx across many different channels. However, these practices are very time consuming since they require massive lab measurements. In this paper, we propose machine learning techniques to cluster post-silicon data from different channels and feed those clusters to Gaussian process regression (GPR) models. We then optimize each GPR surrogate to obtain the optimal tuning settings for each identified cluster. Our methodology is validated by using MATLAB SerDes Toolbox simulations of the functional eye diagram of a Gen6 link.**

***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

The continuously increasing complexity of modern embedded digital systems has made of post-silicon validation a challenging and resource-intensive task. It requires a significant investment of time and computing power, as well as the expertise of many experienced engineers [1]. High-speed input/output (HSIO) links, such as those based on the peripheral component interconnect express (PCIe) standard, constitute a significant portion of circuits to be validated in modern microprocessors. Various optimization methods, including direct and surrogate-based approaches, including space mapping, have been employed to efficiently tune the transmitter (Tx) and receiver (Rx) equalizers [2]. PCIe, as the primary interface between a host central processing unit (CPU) and the input/output (I/O) devices, is one of the most complex HSIO interfaces. The PCIe Gen6 specification, released in 2021, defines a data rate of 64 GT/s. However, as transmission speeds increase, undesirable channel effects such as reflections, crosstalk, jitter, inter-symbol interference (ISI), etc., become more severe, leading to bit errors [3]. Additionally, PCIe channels are bandwidth-limited by default, causing significant signal attenuation at high frequencies, and resulting in distortion and spreading of the transmitted signal over multiple symbols, exacerbating ISI and leading to bit errors. Signal conditioning is a practical solution to overcome these issues and improve the eye diagram [4].

The PCIe specification defines an adaptive equalization (EQ) mechanism to determine the optimal values for the Tx and Rx equalization (EQ) coefficients within a fixed time limit, considering different channel types. The currently most widely used method involves using maps of EQ coefficients obtained from extensive eye diagram measurements. These maps characterize the PCIe link across various channel losses and devices. However, the complete characterization process is very time-consuming and highly dependent on the expertise of validation engineers to select the best EQ values, making it prone to human errors [5].

Machine learning (ML) algorithms offer valuable tools for building statistical models and making predictions based on examples [6]. Unsupervised ML techniques can learn patterns from untagged data, while supervised models are trained to predict outputs based on inputs. In particular, ML techniques can be used to identify patterns based on the analogue behaviour of HSIO links by exploiting the large volume of data generated from post-silicon testing, as we did in [7] for PCIe Gen 5.

In this paper, we extend our ML-based approach in [7] to optimize the new PCIe Gen6 equalization process, which implies reformulating the corresponding objective function to fit the PAM4 characteristic features. Since there are not yet silicon samples with PCIe Gen6, we are validating the proposed method by using MATLAB SerDes Toolbox. We first use unsupervised ML to cluster all available data from different channels. This allows us to classify the data into distinct sets of channel conditions. We then use supervised ML to develop statistical models based on Gaussian process regression (GPR) to predict the eye diagram functional margins within each data subset. Finally, we optimize the GPR-based models to obtain the optimal tuning settings for each of the identified channel conditions.

## II. PCI Express Gen6

PCIe started in 2003 with Gen1 having a data rate of 2.5 giga-transfers per second (GT/s) and has advanced over the years to meet the requirements across different computing markets. After the release of Gen 4, the needs of new applications

demanded faster data transfer, leading to the release in 2019 of Gen5 specification, with a data rate of 32 GT/s [7]. The bandwidth demand from new applications such as artificial intelligence, machine learning, gaming, visual computing, storage, graphics accelerators, high-end networking, coherent interconnects, internet of things (IoT), and memory expanders led to the development of the new Gen6 specification in 2021, reaching data rates of 64 GT/s and adopting PAM4 as the modulation scheme [8].

The PAM4 signaling has four voltage levels, which encodes two bits per voltage level, and uses Gray coding which combines the most significant bit (MSB) and least significant bit (LSB) pairs in a data stream into one of the four voltage levels. By encoding two bits into one symbol, PAM4 achieves the same data rate as NRZ signaling but using half of the bandwidth.

However, while PAM4 solves the bandwidth limitations in high-speed interconnects, it brings new challenges for the physical channel analysis and design. PAM4 has 4 levels and three eye-diagrams, in contrast to the single eye diagram used in PCIe Gen 5 [7]. PAM4 also has a much higher bit error rate (BER) performance, several orders of magnitude higher than the standard $10^{-12}$ BER of the previous PCIe generations. PAM4 also introduces new challenges in slicers, transition jitter, and equalizers, resulting the equalization a critical process with PAM4 signaling [8], for which advanced techniques such as ML are suitable to improve the equalization process.

## III. Unsupervised and Supervised Machine Learning Techniques

Unsupervised ML methods learn patterns from untagged data. As an unsupervised ML approach, clustering aims to find subsets of data based on similarity [9]. Clustering algorithms are divided based on how similarity is measured: they are classified into hierarchical and partitional methods. Partitional methods, particularly, employ distance as means to divide the dataset into clusters.

K-means is a partitional clustering method that iteratively establishes the centroid coordinates of $k$ predefined number of clusters and measures the Euclidean distance between each data point and each centroid. Then, it groups data points based on minimum distance to the cluster centroids [10],[11]. In our work, we employ k-means to cluster margin data points based on channel loss similarities.

Supervised ML aims to classify or predict outcomes based on labelled data. Regression methods look to predict numerical values based on input features. GPR is a non-parametric Bayesian approach to regression. It aims to infer a probability distribution over all possible output values to the model, as opposed to many popular supervised ML algorithms that typically aim to learn exact outputs for a given set of inputs [12]. The non-parametric nature of GPR allows it to calculate the distribution over admissible functions that fit the data, rather than calculating a distribution of parameters for a specific function. This is extremely useful to model output values that inherently have statistical variations, such as physical measurements obtained over varying operating or environmental conditions [14], *e.g.*, post-silicon eye diagram margins. In this work, we use GPR to create non-parametric, probabilistic models based on PCIe6 eye diagram margins data.

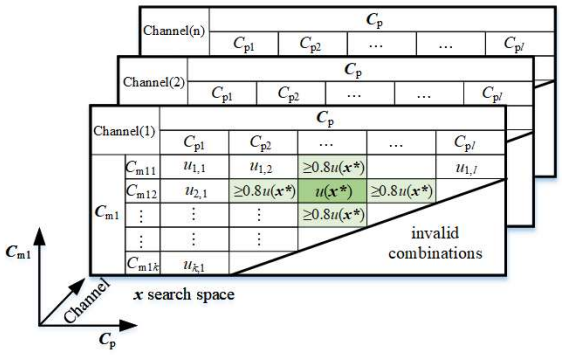


Fig. 1. EQ map coefficients search space for optimization. From [4].

## IV. Transmitter Equalization, Objective Function Formulation and Optimization

### A. Transmitter Equalization

Tx equalization schemes (de-emphasis and pre-emphasis) continue to be used for PCIe Gen6 to open the eye diagram. The Tx equalization coefficients for 64 GT/s are based on a feed-forward equalizer (FFE) 4-tap finite impulse response (FIR) filter. The filter response can be then adjusted by controlling the tap coefficients values, as in [13]. The PCIe specification defines some predefined set of values for the Tx coefficients, referred to as presets. These presets are then adaptively changed during the channel training [8].

However, this adaptive equalization during the computer booting process must be completed within a fixed time limit. Therefore, to reduce the selection time, the current practice is to find out a subset of coefficient combinations during post-silicon validation, and then program it into the system BIOS. Then, during the adaptive equalization process, instead of sweeping all possible values of the coefficients, the adaptive equalization process uses only the subset of coefficients saved into the BIOS [8].

Per the PCIe specification, the values of the Tx coefficients are subjected to some protocol constraints. When all the PCIe specification constraints are applied, the resulting coefficients space may be mapped onto a triangular matrix [4],[13], as shown in Fig. 1, where several EQ maps, one per channel, are superimposed. $C_{m1}$ and $C_p$ coefficients are mapped onto the y-axis and x-axis, respectively. Each matrix cell corresponds to a valid combination of Tx coefficients, and $u(\boldsymbol{x}^*)$ correspond to a combination of $C_{m1}$, $C_p$ that results in a figure of merit (FOM) qualified as optimum.

Per the PCIe specification, the method to find the best subset of coefficients consists of using these EQ maps, which are obtained by measuring the eye diagram characteristics as FOM (i.e. eye height, eye width, or eye diagram area) of the received

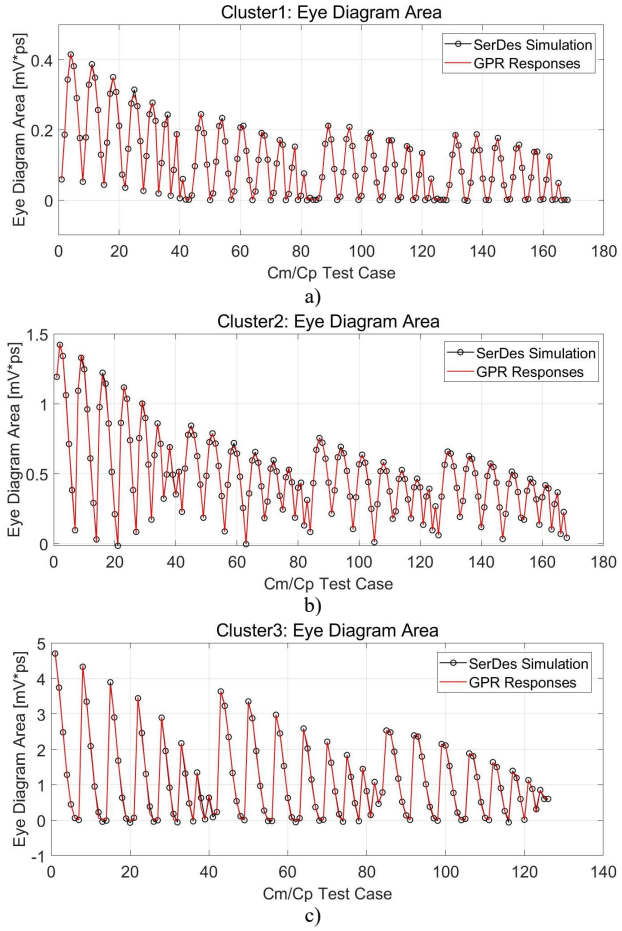


Fig. 2. Comparing the resultant GPR models responses versus Gen6 Link SerDes Simulation across all possible channels and $C_m/C_p$ combinations: a) cluster #1 (hi-loss), b) cluster #2 (mid-loss), c) cluster #3 (low-loss).

signal for each of the $C_{m1}$, and $C_p$ combinations across each channel and device pairing, requiring multiple EQ maps. The method consists of finding the set of Tx coefficients that qualify the FOM as near optimal. Due to the large number of EQ maps, along with all the new challenges imposed by PAM4, finding the optimal subset of coefficients would be a very challenging task, and then an optimization procedure may be used to confront this challenge.

### *B. Objective Function Formulation and Optimization*

We aim at finding the optimal set of Tx EQ settings to maximize the eye width ($e_w$) and eye height ($e_h$). These two eye diagram parameters are the smallest of the three PAM4 eye height measurements (lower, middle, and upper), which are functions of the Tx FFE EQ settings contained in vector $\boldsymbol{x}$, *i.e.*,

$$e_w(\boldsymbol{x}) = \min(e_{w\text{-low}}(\boldsymbol{x}), e_{w\text{-mid}}(\boldsymbol{x}), e_{w\text{-upp}}(\boldsymbol{x})) \quad (4)$$

$$e_h(\boldsymbol{x}) = \min(e_{h\text{-low}}(\boldsymbol{x}), e_{h\text{-mid}}(\boldsymbol{x}), e_{h\text{-upp}}(\boldsymbol{x})) \quad (5)$$

We follow [4] to define an initial objective function as

$$u(\boldsymbol{x}) = -[e_w(\boldsymbol{x})][e_h(\boldsymbol{x})] \quad (6)$$

$e_w$ and $e_h$ are functions of the Tx finite impulse response (FIR) filter pre-cursor ($C_{m1}$) and post-cursor ($C_p$) EQ coefficient values (integer numbers), contained in vector $\boldsymbol{x}$.

We are looking to do a SBO in a single EQ map, but we may have a large number of EQ maps depending on the number of channels. Then, we compute the median of all $u(\boldsymbol{x})$ values across all channels. 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 $u(\boldsymbol{x}^*)$ [7].

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 [7],

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} u(\boldsymbol{x}) \quad (7)$$

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

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

$$\boldsymbol{l}(\boldsymbol{x}) = 0.8u(\boldsymbol{x})\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} - \begin{bmatrix}u(\boldsymbol{x}-[1\ 0]^{\mathrm{T}}) & u(\boldsymbol{x}+[1\ 0]^{\mathrm{T}})\\ u(\boldsymbol{x}-[0\ 1]^{\mathrm{T}}) & u(\boldsymbol{x}+[0\ 1]^{\mathrm{T}})\end{bmatrix} \quad (8)$$

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

$$U(\boldsymbol{x}) = -[e_w(\boldsymbol{x})][e_h(\boldsymbol{x})] + L(\boldsymbol{x})\left[\frac{|u(\boldsymbol{x}^{(0)})|}{\max\{\boldsymbol{l}(\boldsymbol{x}^{(0)})\}}\right] \quad (9)$$

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

$$L(\boldsymbol{x}) = \max\{0, \max\{\boldsymbol{l}(\boldsymbol{x})\}\} \quad (10)$$

Our unconstrained objective reduces to

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} U(\boldsymbol{x}) \quad (11)$$

We aim at finding the optimal set of EQ coefficients values $\boldsymbol{x}^*$ by solving (11) using the gradient-free computationally efficient direct-search optimization method Nelder-Mead [16].

## V. Results

Because there are not still silicon samples with PCIe Gen6, we use MATLAB SerDes Toolbox to validate our methodology considering a short, medium, and long-reach channels (CEI-56G serial links) of 1dB to 5dB, 9dB to 15dB, and 20dB to 27dB losses, respectively, in a 64 GT/s PCIe Gen6 link, where the pass/fail criteria is defined in terms of a time domain eye diagram at BER=$10^{-6}$. The link is simulated with the corresponding Tx jitter parameters (deterministic and sinusoidal) based on the PCIe Specification, and Rx jitter parameters from a common reference clock Rx architecture. The simulator generates an output containing the three statistical eye heights and widths.

We first cluster all available eye diagram margins across different channels, dividing them into distinct sets of channel losses. 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.

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. A comparison between

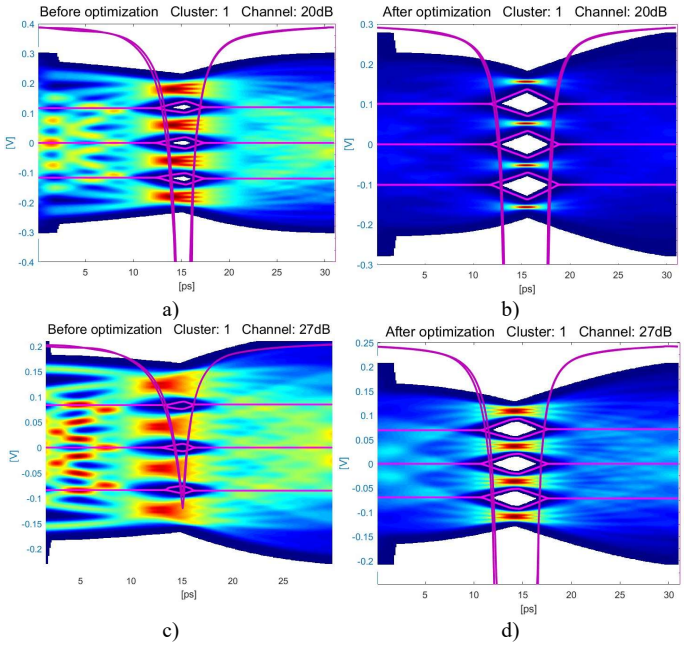


Fig. 3. Eye diagram before and after the optimization process: a), b): 20dB losses channel; c), d): 27dB losses channel.

the resultant GPR models' responses and SerDes simulations are shown in Fig. 2, where an excellent overall match is seen. We evaluate the accuracy of the obtained GPR models by using an average relative error defined as in [15]. The relative error for cluster 1 is 3.8%, for cluster 2 is 2.8%, and for cluster 3 is 5.9%. As shown in Fig. 2, the resultant GPR models are able to follow the overall trends of the actual measurements. Therefore, these models are quite suitable to perform effective and inexpensive surrogate-based optimization (SBO).

We next perform a SBO with these GPR models using the formulation in Section IV-B, to obtain the optimal tuning Tx equalizer settings at a low computational cost. We validate the SBO results by measuring (from simulations) the PCIe Gen6 link Rx inner eye height/width at $\boldsymbol{x}^*$. Figure 3 shows the eye diagrams at the receiver, before and after applying the optimization process. The results confirms that the resultant top eye width and height meets the channel tolerancing eye mask defined in the PCIe Gen6 Spec even at worst-case condition of jitter and channel losses.

## VI. Conclusion

We proposed a ML technique to cluster post-silicon data from different PCIe Gen6 link channels and feed those clusters to GPR models. We then used an SBO approach for 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. Our methodology is validated by using MATLAB SerDes Toolbox simulations of the functional eye diagram of a Gen6 link. The optimized EQ coefficients were validated by measuring the simulated eye diagram margins at the receiver, demonstrating a significant increase in eye diagram margins. This methodology can easily be applied on post-silicon industrial validation once the first processors with PCIe Gen6 arrive to the validation site, allowing to accelerate 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] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*. London, UK: Elsevier Academic Press. 2015.

[7] E. Rangel-Patiño, A. Viveros-Wacher, C. Rajyaguru, E. A. Vega-Ochoa, S. D. Rodriguez-Saenz, J. L. Silva-Cortes, H. Shival, and J. E. Rayas-Sánchez, "PCIe Gen5 physical layer equalization tuning by using K-means clustering and Gaussian process regression modeling in industrial post-silicon validation," in *IEEE MTT-S Int. Conf. Numer. EM Mutiphysics Modeling Opt. (NEMO-2023)*, Winnipeg, Canada, Jun. 2023, pp. 126-129.

[8] 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.*, Cali, Colombia, May 2021, pp. 1-4.

[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] K. G. López-Araiza, F. E. Rangel-Patiño, J. E. Ascencio-Blancarte E. A. Vega-Ochoa, J. E. Rayas-Sánchez, and O. Longoria-Gándara, "A multi-stage CTLE design and optimization for PCI Express Gen6.0 link equalization," in *IEEE Latin American Electron Devices Conf. (LAEDC)*, Puebla, Mexico, Jul. 2023, pp. 1-4.

[14] 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.

[15] 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.

[16] 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.