Eye Diagram System Margining Surrogate-Based Optimization in a Server Silicon Validation Platform

Francisco E. Rangel-Patiño[1,2], José L. Chávez-Hurtado[2], Andres Viveros-Wacher[1,2], José E. Rayas-Sánchez[2], and Nagib Hakim[3]

[1]*Intel Corp., Zapopan, Jalisco, 45019 Mexico*

[2]*Department of Electronics, Systems, and Informatics, ITESO – The Jesuit University of Guadalajara, Tlaquepaque, Jalisco, 45604 Mexico*

[3]*Intel Corp., Santa Clara, CA, 95052 USA*

erayas@iteso.mx

***Abstract*— Exhaustive enumeration methods for the physical layer (PHY) tuning of high-speed input/output (HSIO) links are prohibitive under current silicon server time-to-market (TTM) commitments. An alternative is to perform optimization on a highly accurate surrogate model. However, to increase the accuracy of the model, the number of lab measurements required to derive it also increases. In this paper, we analyze several surrogate modeling methods and design of experiments techniques to find the coarse model that is capable of approximating the real system behavior without requiring a large amount of actual measurements. We perform a direct optimization on the best coarse models found and verify the response by measuring the real system at the optimal coarse model solution.**

***Index Terms* — DoE, equalization, eye diagram, high-speed interconnects, HSIO, Kriging, neural network, optimization, polynomial, post-silicon validation, receiver, signal integrity, support vector machines, surrogate models, USB3.**

I. INTRODUCTION

The combined effects of increased product complexity, performance requirements, and time-to-market (TTM) commitments have added tremendous pressure on post-silicon validation [1]. Within the computer server segment, there are conditions that further increase system complexities. These include increased I/O density, decreased power consumption, as well as non-flexible form factors [2]. The latter implies that channel designs remain unchanged, thus turning the problem towards analog circuitry optimization. Therefore, physical layer (PHY) tuning based on equalization techniques are used to cancel any undesired effect such as transmitter jitter, attenuation or inter-symbol interference, among others [3], [4]. The current industrial practices to perform PHY tuning consist of an exhaustive enumeration method, turning them into the most time-consuming processes in post-silicon validation [1], [5], [6]. To perform PHY tuning, the receiver (Rx) eye diagram margins [7] are optimized until comply with the link specifications. Accurate direct simulations for PHY tuning in high-speed input/output (HSIO) links are computationally very expensive given the complexity of the system involved. On the other hand, surrogate models are scalable mathematical models that can be used as a parameterized approximation of a system response within a design space of interest [8], [9]. While an accurate surrogate model is desirable for direct surrogate-based optimization (SBO), it can be very expensive to derive it, since it requires massive lab measurements which are prohibitive under the current TTM schedules. However, by combining a good modeling technique with a suitable design of experiments (DoE) approach, an efficient surrogate model can be developed.

In this paper, we develop surrogate models of a HSIO link based on actual measurements of an industrial server post-silicon validation platform. We compare several surrogate modeling techniques combined with different DoE approaches to find the best coarse model, verifying the response of the resultant coarse models by comparing with actual measurements. We next perform a surrogate-based optimization (SBO) with the best coarse models found to obtain the optimal PHY tuning Rx equalizer settings. We finally validate our approach by measuring the actual functional eye diagram on the real system using the optimal settings predicted by the coarse model.

This paper uses the same methodology proposed in [10] but in a different scenario: a) we consider a different industrial HSIO link; b) we find a suitable combination of DoE and surrogate technique to develop an effective coarse model with a very reduced set of data (for a future space mapping optimization [11]), in contrast to [10] where we aim at finding the most accurate surrogate model at the expense of a higher number of measurements. The paper is organized as follows: Section II describes the measurement setup. The SBO, DOE approaches, and modeling formulation are presented in Section III; results are discussed in Section IV. Finally, Section V concludes our work.

II. SYSTEM DESCRIPTION

The system under test is an Intel server post-silicon validation platform in an industrial environment, as shown in Fig. 1. The platform is comprised mainly of a CPU and a platform controller hub (PCH) [12]. Within the PCH, our methodology was tested on a USB3.1 Gen1 HSIO link [13].

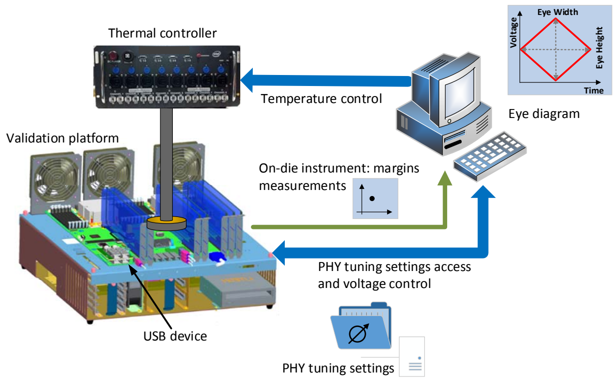

Fig. 1. Test setup: an Intel server post-silicon validation platform.

The measurement system is based on system margin validation (SMV) [14], which is a methodology that consists of measuring the Rx functional eye width and eye height by using on-die design for test (DFT) features until the eye opening has been shrunk to a point where the Rx detects errors or the system fails [10].

III. Surrogate-Based Optimization

A large amount of data is usually needed to ensure surrogate model accuracy. However, generating large amounts of data is very expensive in the post-silicon validation environment. DoE can be exploited to reduce the dimension of these data sets, ensuring adequate parameter coverage [15]. Then surrogate models can be constructed using this data from measurements, and provide fast approximations of the original system at new design points [16]. In this section, several surrogate modeling techniques combined with several DoE approaches are explored to construct an efficient surrogate model for PHY equalizer simulation.

A. Design of Experiments Approaches

The accuracy of surrogate models depends on the amount of training and testing data, at the expense of a higher number of evaluations in the real validation environment. Therefore, we employ three different DoE techniques to explore the desired solution space with a far less number of runs: Box Behnken (BB), orthogonal arrays (OA), and Sobol. For each technique, the input variables represent Rx PHY parameters; in this case we use 8 variables and 3 levels for each variable. We then retrieve the eye measurements from the system under test. The samples taken are later used as the training and testing data required for surrogate modeling [10].

BB is a type of second order RSM design that combines factorial designs with balanced incomplete blocks designs [17]. With 8 variables and 3 levels for each variable, the number of experiments in BB is 120.

OAs are experimental designs identified by $L_N(s^k)$, where N is the number of experimental runs, s is the number of states (or levels) for each variable, and k is the number of variables [18]. We use an $L_{27}(3^8)$ OA in our work yielding only 27 experiments.

We select the Sobol [19] low-discrepancy sequence as the third DoE option to sample the solution space. Given the quasi-Monte Carlo sampling approach of 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 two different Sobol DoE, denoted as Sobol50 and Sobol100, with 50 and 100 samples, respectively.

B. Surrogate Modeling Formulation

In this work, the purpose of the surrogate models is to represent, as accurately as possible, the behavior of the electrical margining system, which consist of the eye height e_{h} and eye width e_{w} of the measured functional eye diagram.

The electrical margining system response, denoted as $\boldsymbol{R}_{\mathrm{f}}$, depends on the Rx PHY tuning parameters $\boldsymbol{x}$, the system operating conditions $\boldsymbol{\psi}$, and the devices under test $\boldsymbol{\delta}$.

Then, the response of the surrogate models $\boldsymbol{R}_{\mathrm{s}}$ must be as close as possible to the original fine model response $\boldsymbol{R}_{\mathrm{f}}$,

$$\boldsymbol{R}_{\mathrm{s}}(\boldsymbol{x},\boldsymbol{\psi},\boldsymbol{\delta}) \approx \boldsymbol{R}_{\mathrm{f}}(\boldsymbol{x},\boldsymbol{\psi},\boldsymbol{\delta}) \tag{1}$$

C. Surrogate Modeling Techniques

Five different techniques are implemented in this work to develop a surrogate model for the electrical margining system [10]: a) polynomial-based surrogate modeling (PSM), b) support vector machines (SVM), c) Kriging, d) generalized regression neural networks (GRNN), and e) a three-layer perceptron neural network (3LPANN).

PSM is a surrogate modeling technique that exploits the multinomial theorem to represent the system behavior using polynomials [20]. GRNN is a special type of artificial neural networks (ANN) where the number of hidden neurons is equal to the number of learning samples [21] and does not require an iterative training procedure [22]. SVM exhibits a good tradeoff between model complexity and generalization capability by exploiting kernel functions to find the optimal model parameters [23]. Kriging technique aims to minimize the prediction variance by exploiting the best linear unbiased estimator of the output value for a given input [24]. Finally, a three-layer perceptron neural network is implemented. The number of hidden neurons h increases (starting with $h = 1$) until generalization error deteriorates and the current learning error is smaller than the current testing error [25].

PSM was implemented in Matlab following [20], while the other four surrogate model techniques were implemented using the corresponding Matlab Toolboxes with default parameters.

D. Surrogate-Based Optimization

A direct optimization algorithm is applied to the selected surrogate model to find the optimal parameters for the PHY tuning settings. We are looking to maximize the corresponding eye diagram by minimizing the objective function [10]

TABLE I
SURROGATE MODELS GENERALIZATION ERROR FOR EYE HEIGHT

model	BB	OAL27	Sobol50	Sobol100
PSM	6.07%	29.60%	4.73%	5.70%
SVM	5.47%	42.18%	5.09%	4.65%
Kriging	5.91%	41.77%	5.42%	4.63%
GRNN	14.77%	45.44%	14.22%	10.13%
3LPANN	7.32%	40.77%	5.74%	6.19%

TABLE II
SURROGATE MODELS GENERALIZATION ERROR FOR EYE WIDTH

model	BB	OAL27	Sobol50	Sobol100
PSM	11.61%	21.56%	18.97%	8.24%
SVM	20.15%	21.65%	21.00%	19.04%
Kriging	16.48%	21.06%	15.83%	13.61%
GRNN	20.46%	26.79%	22.06%	18.04%
3LPANN	12.28%	19.13%	6.73%	7.35%

$$u(\boldsymbol{x}) = -[e_{\mathrm{w}}(\boldsymbol{x},\boldsymbol{\psi},\boldsymbol{\delta})][e_{\mathrm{h}}(\boldsymbol{x},\boldsymbol{\psi},\boldsymbol{\delta})] \tag{2}$$

The optimal PHY tuning settings are found by solving

$$\boldsymbol{x}^{*} = \arg\min_{\boldsymbol{x}} u(\boldsymbol{x}) \tag{3}$$

Once a set of optimal parameters is found, it is tested on the actual electrical margining system to corroborate the eye diagram area has been maximized.

IV. RESULTS

We first evaluate the accuracy of the obtained surrogate models by comparing with actual measured responses from a USB3.1 Gen1 HSIO link. In all the modeling cases, we use 150 samples testing base points not seen during training to measure the generalization performance.

Tables I and II show a summary of the generalization performance for the eye height and eye width, comparing the five surrogate models using four DoE: a) OAL27, b) BB, c) Sobol50, and d) Sobol100. It is seen from those tables that, overall, the PSM and the 3LPANN techniques yield the lowest testing average relative errors for both eye measurements when using Sobol100, which is the DoE technique yielding the best generalization performance at the expense of the highest number of measurements on the real system.

Testing the Rx functional eye diagram is a time-consuming task taking around 38 minutes for a single test case on a USB3.1 HSIO link. Therefore collecting data to build a model from a Sobol100 or BB can take up 76 hours, which is excessive in the post-silicon validation environment. On the other hand, Tables I and II show Sobol50 and OAL27 to be DoE approaches that can yield good-enough surrogate models.

Fig. 2 shows the absolute error at all testing samples for both eye height and eye width, for the five surrogate models using OAL27 and Sobol50 DoE techniques. When using OAL27 DoE, it is observed that the PSM model shows a good accuracy (Fig. 2a, 2c), while the 3LPANN model with Sobol50 provides also a good accuracy (Fig. 2b, 2d). From here, we take PSM with OAL27 and 3LPANN with Sobol50 as the candidates to be our coarse models.

We next perform a SBO as described in Section III.D, with the two coarse model candidates to obtain near optimal PHY tuning Rx equalizer settings. Finally, we validate the SBO results by measuring the USB3.1 link Rx inner eye height/width at $\boldsymbol{x}^{*}$ on the real validation platform with a commercial USB device, and compare the SBO results with the exhaustive enumeration method. The results, shown in Fig. 3, indicate that our approach yields better results in terms of eye diagram area and development time than the exhaustive method. Differences in eye diagram between PSM with

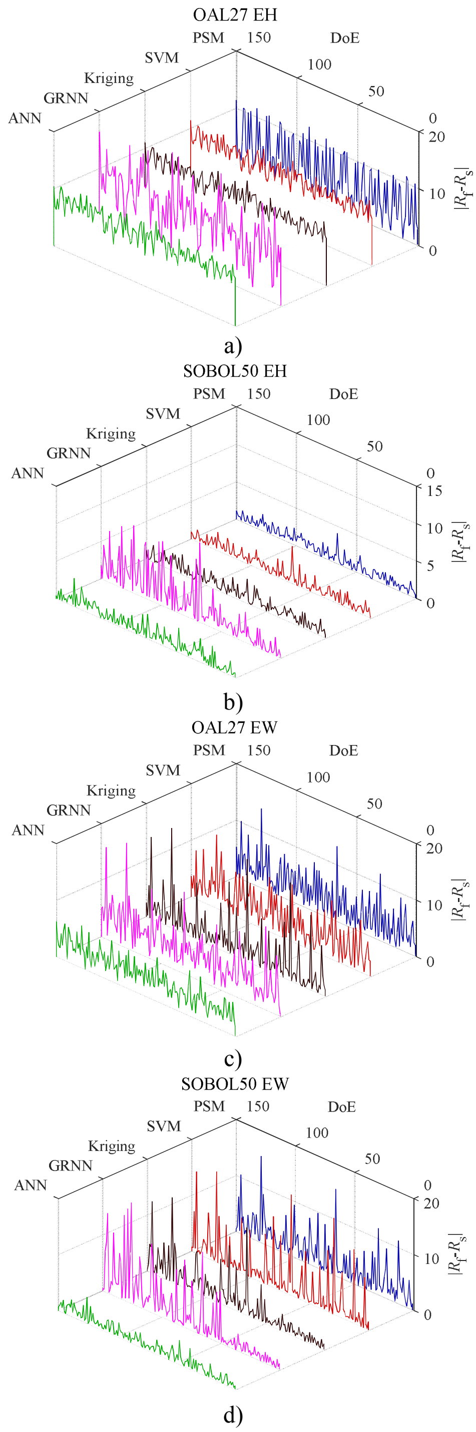

Fig. 2. Surrogate models absolute testing errors: a) OAL27 and b) Sobol50, for eye height; c) OAL27 and d) Sobol50, for eye width.

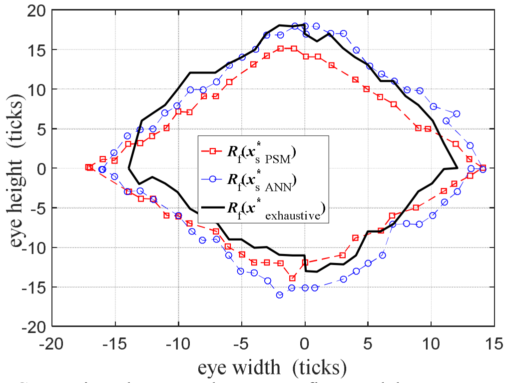

Fig. 3. Comparison between the system fine model responses after SBO using PSM with OAL27, and 3LPANN with Sobol50 surrogate models versus exhaustive enumeration method.

OAL27 and 3LPANN with Sobol50 are not very significant. However, there is almost a 50% of time savings (around 16 hours) by using OAL27 as compared against Sobol50. Therefore, using PSM modeling technique with OAL27 as the DoE approach yields an efficient coarse model of the system.

V. Conclusions

We presented several surrogate models trained with different DOE techniques to find a good coarse model able to approximate a real server HSIO link with a very reduced amount of measurements. We selected the best combinations of surrogate modeling technique and DOE in terms of accuracy and development time. Then we applied direct optimization on the best coarse models found and compare their performance by measuring the real system at the optimal coarse model solutions. Through this procedure, we found an efficient coarse model that approximates the system with a very reduced set of testing and training data, suitable for a future space mapping optimization.

References

[1] J. Keshava, N. Hakim, and C. Prudvi, "Post-silicon validation challenges: How EDA and academia can help," in *Proc. Design Automation Conference (DAC)*, Anaheim, CA, June 2010, pp. 3-7.

[2] B. T. Lee, M. Mazumder, and R. Mellitz, "High speed differential I/O overview and design challenges on Intel enterprise server platforms," in *IEEE Int. Symp. on Electromagnetic Compatibility*, Long Beach, CA, Aug. 2011, pp. 779-784.

[3] W. Hodgkiss and L. F. Turner, "Practical equalization and synchronization strategies for use in serial data transmission over H.F. channels," *Radio and Electronic Engineer*, vol. 53, no. 4, pp. 141-145, April 1983.

[4] H. Zhang, S. Krooswyk, and J. Ou, *High Speed Digital Design: Design of High Speed Interconnects and Signaling*. Waltham, MA: Elsevier, 2015.

[5] X. Wang and Q. Hu, "Analysis and optimization of combined equalizer for high speed serial link," in *Proc. IEEE Int. Conf. Anti-counterfeiting, Security, Identification (ASID)*, Xiamen, China, Sept. 2015, pp. 43-46.

[6] Y. S. Cheng and R. B. Wu, "Direct eye diagram optimization for arbitrary transmission lines using FIR filter," *IEEE Trans. Components, Packaging Manufact. Techn.*, vol. 1, no. 8, pp. 1250-1258, Aug. 2011.

[7] A. Viveros-Wacher, R. Alejos, L. Alvarez, I. Diaz-Castro, B. Marcial, G. Motola-Acuna, and E. A. Vega-Ochoa, "SMV methodology enhancements for high speed IO links of SoCs," in *IEEE VLSI Test Symp. (VTS-2014)*, Napa, CA, May. 2014, pp. 1-5.

[8] M. B. Yelten, T. Zhu, S. Koziel, P. D. Franzon, and M. B. Steer, "Demystifying surrogate modeling for circuits and systems," *IEEE Circuits Systems Magazine*, vol. 12, pp. 45-63, First Quarter 2012.

[9] O. Garitselov, S. P. Mohanty, and E. Kougianos, "A comparative study of metamodels for fast and accurate simulation of nano-CMOS circuits," *IEEE Trans. Semicond. Manufac.*, vol. 25, no.1, pp. 26-36, Feb. 2012.

[10] 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. Microwave Theory Techn.*, vol. 65 (work submitted to LAMC-2016 Mini-Special Issue).

[11] J. E. Rayas-Sánchez, "Power in simplicity with ASM: tracing the aggressive space mapping algorithm over two decades of development and engineering applications," *IEEE Microwave Magazine*, vol. 17, no. 4, pp. 64-76, Apr. 2016.

[12] F. Rangel-Patino, A. Viveros-Wacher, J. E. Rayas-Sanchez, 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)*, Puerto Vallarta, Mexico, Dec. 2016, pp. 1-3.

[13] USB Org. (2016). *Universal Serial Bus Revision 3.1 Specification* [Online]. Available: http://www.usb.org/developers/doc

[14] C. Gu, "Challenges in post-Silicon validation of high-speed I/O links," in *Proc. IEEE/ACM Int. Conf. on Computer-Aided Design (ICCAD)*, San Jose, CA, Nov. 2012, pp. 547-550.

[15] Y. Mack, T. Goel, W. Shyy, and R. Haftka, "Surrogate Model-Based Optimization Framework: A Case Study in Aerospace Design," *Studies in Computational Intelligence*, vol. 51, no. 2, pp. 323-342, 2007.

[16] N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathna, and P. K. Tucker, "Surrogate-based analysis and optimization," *Prog. in Aerospace Sciences*, vol. 41, no. 1, pp. 1-28, Jan. 2005.

[17] C. F. J. Wu and M. Hamada, *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York, NY: Wiley, 2000.

[18] Y. P. Chang et al, "Design of discrete-value passive harmonic filters using sequential neural-network approximation and orthogonal array", in *IEEE PES Transmission and Distribution Conf. & Expo.: Asia and Pacific*, Dalian, China, Aug. 2005, pp. 1-6.

[19] 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. 1, pp. 86-112, Jan. 1967.

[20] J. L. Chávez-Hurtado and J. E. Rayas-Sánchez, "Polynomial-based surrogate modeling of RF and microwave circuits in frequency domain exploiting the multinomial theorem," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 12, pp. 4371-4381, Dec. 2016.

[21] P. Mahouti, F. Günes, S. Demirel, A. Uluslu, and M. A. Belen, "Efficient scattering parameter modeling of a microwave transistor using generalized regression neural network," in *IEEE Int. Conf. on Microwaves, Radar, and Wireless Communication (MIKON)*, Gdansk, Poland, Jun. 2014, pp. 1-4.

[22] D. F. Specht, "A general regression neural network," *IEEE Trans. on Neural Networks*, vol. 2, no. 6, pp. 568-576, 1991.

[23] N. T. Tokan and F. Günes, "Analysis and synthesis of the microstrip lines based on support vector regression," in *Proc. IEEE Eur. Microw. Conf. (EuMC)*, Amsterdam, The Netherlands, Oct. 2008, pp. 1473-1476.

[24] W. C. M. Van Beers, "Kriging metamodeling in discrete-event simulation: An overview," in *Proc. Simulation Conf.*, Dec. 2005, pp. 202-208.

[25] 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 Techn.*, vol. 54, no. 12, pp. 4528-4537, Dec. 2006.