Fast Design Optimization through Simple Kriging Metamodeling: A Sense Amplifier Case Study

Oghenekarho Okobiah, Student Member, IEEE, Saraju P. Mohanty, Senior Member, IEEE, Elias Kougianos, Senior Member, IEEE

***Abstract*—With the increasing complexity of nanoscale-CMOS circuits and systems integration, full SPICE simulations for silicon accurate results have run times in the order of days or weeks. This brief presents a methodology that uses a simple Kriging metamodeling technique capable of modeling the correlation effects between parameters, and a simulated annealing algorithm for ultra-fast design optimization. The proposed methodology is applied to a clamped bitline amplifier circuit and shows promising results for increased accuracy in process-aware metamodeling techniques. The error of the metamodels is very small and they are generated in 10.5 mins compared to the 72 hours taken for an exhaustive simulation. The design optimization performed on the metamodels improved the precharge time of the circuit by 61.15%.**

***Index Terms*—Kriging Methods, Metamodeling, Fast Design Optimization, Nano-CMOS Analog Circuits**

I. Introduction

Full simulation of complex nanoscale circuits takes days and sometimes weeks. With a large number of parameters and pronounced effects from process variation, efficient and accurate design optimization becomes thus prohibitive. The inherent problem is one of computational efficiency versus accuracy. Metamodeling techniques based on low-order polynomial regression are one of the most common methods used [2] to reduce run times. In regression based techniques, the errors due to process variation are assumed to be random and uncorrelated and are thus equally approximated across the design space, which leads to inaccurate models for global optimization. In the case of many nano-CMOS designs, where the effects of process variation are a significant factor, the errors are usually correlated among design parameters. Kriging methods use a combination of global trend functions and local departures to create more accurate models [3]. The local departure function is a correlation function which accurately models the nanoscale effects making Kriging metamodels process aware and robust for high dimensional designs. Metamodeling based optimization design using Kriging prediction techniques has been explored in many fields but only recently in VLSI [4], [5]. In [4], Kriging methods are used to iteratively extract Pareto models. In [5], Kriging models are generated for an operational amplifier.

O. Okobiah and S. P. Mohanty are with Computer Science and Engineering, University of North Texas, Denton, Email-IDs: oo0032@unt.edu, saraju.mohanty@unt.edu. E. Kougianos is with Engineering Technology, University of North Texas, Denton, E-mail: elias.kougianos@unt.edu. This research is supported in part by NSF award DUE-0942629. A preliminary version of this research is presented at: [1].

The **novel contribution of this paper** is a methodology that combines metamodels and optimization algorithms for fast design optimization of analog/mixed-signal circuits. The speedup comes from the use of metamodels (instead of SPICE netlist), automatic optimization algorithm, and elimination of multiple manual layout steps. The metamodels are then optimized using a simulated annealing based algorithm. Specifically, the contributions of this paper are: 1) A fast methodology combining metamodeling with automatic optimization algorithms is presented which reduces analog design cycle significantly. 2) Simple Kriging metamodeling is explored and its accuracy analysis with "golden" data is performed. 3) A case study circuit for a 45 nm CMOS based clamped bitline sense-amplifier (CBLSA) is presented. 4) A sense amplifier optimization algorithm that uses the Kriging metamodels and converges to a solution in reasonable time is proposed.

The rest of this paper is organized as follows: The simple Kriging method is introduced in Section II. The proposed fast flow is discussed in Section III. Experimental results of the proposed methodology using a 45nm CMOS CBLSA as case study are presented in Section IV. Conclusions and future research are discussed in Section V.

II. Fundamentals of Simple Kriging Metamodels

The main idea behind Kriging is that the predicted outputs are weighted averages of sampled data. The weights are unique to each predicted point and are a function of the the distance between the point to be predicted and observed points [6], [7]. The general expression of a Kriging model is as follows:

$$y(\mathbf{x_0}) = \sum_{j=1}^{M} \lambda_j B_j(\mathbf{x}) + z(\mathbf{x}), \tag{1}$$

where $y(\mathbf{x_0})$, is the predicted response at design point $\mathbf{x_0}$ $\{B_j(\mathbf{x}), j = 1, \cdots, M\}$ is a specific set of basic functions over the M-dimensional design domain D_M, λ_j are fitting coefficients (or weights) to be determined and $z(\mathbf{x})$ is the random process error. In simple Kriging, a constant and known mean over the global domain is assumed for the predicted point. It is assumed that the process has a mean μ, variance σ^2, and correlation function, called the "variogram" in geostatistics", $r(\mathbf{s}, \mathbf{t})$ between points $\mathbf{s}$ and $\mathbf{t}$ given by:

$$r(\mathbf{s}, \mathbf{t}) = \mathrm{Corr}(z(\mathbf{s}), z(\mathbf{t})). \tag{2}$$

The variogram is used to derive the Kriging weights, λ_j. The autocorrelation of the design points is characterized by the covariance function [8]. The weights are chosen so that the

Kriging variance is minimized and the weighting scheme is given by:

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \Gamma^{-1} \begin{pmatrix} \gamma(e_1, e_0) \\ \vdots \\ \gamma(e_n, e_0) \\ 1 \end{pmatrix}, \quad (3)$$

where Γ is the covariance matrix of the observed points. One advantage of Kriging is that the estimated response at sample points is exactly the same as the observed data [8].

Estimation of the correlation between sampled points and a predicted point is done with the semivariogram model. Based on the nature of the observed data points, the empirical model could be fit to either spherical, linear, Gaussian or exponential theoretical models. The smoothness of the predicted responses is affected by the theoretical model used. A steeper model reduces the smoothness because it places more weight on closer neighbors. The most common model used is the spherical and with C_0, C and a are shape parameters it is expressed by:

$$\gamma(h) = C_0 + C\left(\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right) \text{ for } 0 < h \leq a. \quad (4)$$

III. Proposed Kriging-based Fast Methodology

Metamodels are used to increase the efficiency of the design optimization while maintaining a sufficient accuracy. Kriging methods, which take into account the correlations among design parameters, are excellent prediction models for complex designs with high dimensionality. The proposed fast design flow is shown in Fig. 1 and is divided into four steps as discussed below. The flow can be modified to incorporate statistical models to accommodate nanoscale effects [9].

A. Parasitic-Aware Netlist Generation and Parameterization

The starting point of the flow is a parasitic-aware netlist extracted from the physical layout design and used in order to achieve silicon-level accuracy. Design (length L and width W) and process (threshold voltage V_{th} and oxide thickness (T_{ox})) parameters are identified in the parasitic-aware netlist which is then parameterized with respect to these variables for automatic sample point generations. The parameterization ensures that the layout design does not have to be physically redrawn during each iteration. The main assumption being that the resizing of the devices does not perturb the interconnect significantly. We call this on-the-fly automatic layout-accurate netlist resizing "virtual resizing".

B. Accurate and Fast Design Space Sampling

Latin Hypercube Sampling (LHS) is used for generating sample design points. LHS covers all input dimensions simultaneously thus improving on the variance compared to random Monte Carlo distributions. The variance of the mean $\overline{y}_{LHS}$ of a function $f(\mathbf{x})$ over n LHS sample points and is given by [10]:

$$Var(\overline{y}_{LHS}) = \frac{1}{n} Var(f(\mathbf{x})) - \frac{k}{n} + o\left(\frac{1}{n}\right), \quad (5)$$

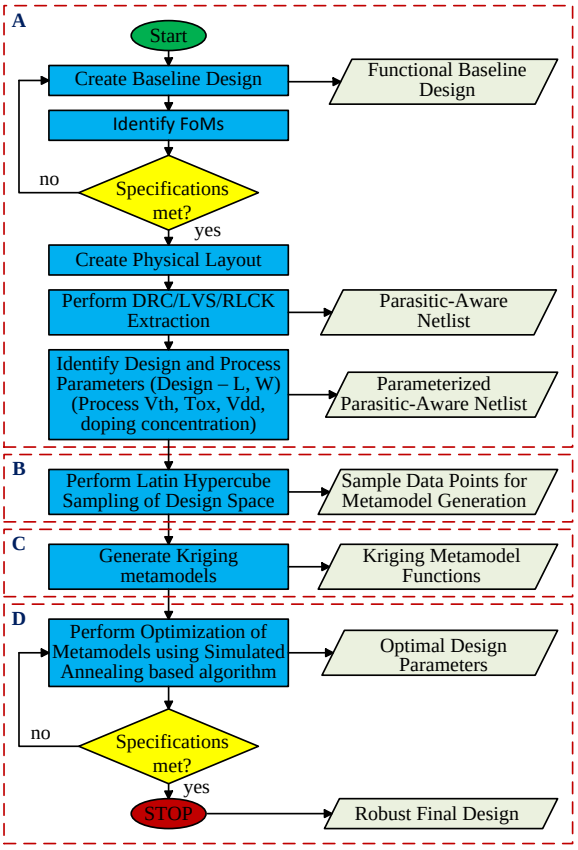


Fig. 1. The proposed Kriging assisted fast and accurate design optimization.

where k is a positive constant shown to be smaller than the variance of random samples. A comparison of sampling techniques and sample sizes have been performed and LHS is preferred [11]. LHS sample point responses are generated using analog simulations and are fed into the Kriging metamodel generator. L and W are used as design parameters while the process parameters are varied to model the effects of process variation.

C. Simple Kriging Metamodel Generation

The generated metamodel is a function of the design parameters L and W, and process parameters. In this paper, metamodels are generated using W_n and W_p. A total of four metamodels are generated; one for each of the Figures-of-Merit (FoMs) in the CBLSA design: precharge time T_{PC}, sense delay T_{SD}, average power P_{SA} and sense margin V_{SM}. Each FoM can be expressed based on the general form of the Kriging function. For example, the predicted precharge time $\widehat{Y}_{pr}$ at an *unknown design point* W_n^* is expressed as follows:

$$\widehat{Y}_{pr}\left(W_n^*\right) = \sum_{i=1}^{N} \lambda\left(W_n^*\right)_i Y_{pr}\left(W_{n_i}\right), \quad (6)$$

where $Y_{pr}(W_{n_i})$ are the observed precharge values for the given N W_{n_i} $(i = 1, 2, \ldots, N)$ sample points. The weights $\lambda(W_n^*)$ are unique for each predicted point W_n^* and are calculated from Eqn. (3). Algorithm 1 summarizes the process of the simple Kriging metamodel generation.

Algorithm 1 Simple Kriging based metamodel generation for various FoMs of the clamped bitline sense amplifier.

1: Obtain the target specifications of the CBLSA design and select the performance objectives or FoMs.
2: Create the parameterized parasitic-aware netlist of the CBLSA circuit after performing the baseline physical design, DRC and LVS verification, and RCLK extraction.
3: Initialize the number of sample set points (n).
4: Generate n sample set points using LHS.
5: Obtain n sample points $D = [D_1, \ldots, D_n]$ for M design variables using LHS.
6: Derive variogram model for each FoM based on the observed sample points.
7: **for** Each design point to be predicted. **do**
8: Generate the variogram for simple Kriging.
9: Generate prediction weights for simple Kriging.
10: Generate simple Kriging models for design points.
11: **end for**
12: Perform accuracy analysis of the simple Kriging metamodels using the Root Mean Square Error (RMSE) and the correlation coefficient R^2.

D. Algorithm for Optimization over Kriging Metamodels

A simulated annealing based algorithm is proposed to optimize the simple Kriging metamodels of the CBLSA. The metamodels can be optimized for each of the identified FoMs. In this paper, the precharge time (T_{PC}) is used as the objective while the average power consumption (P_{SA}) is used as a design constraint. The optimization steps are presented in Algorithm 2. The algorithm used to generate the Kriging metamodels was written using MATLAB with the help of the toolboxes mGstat [12] and SUMO [13].

IV. Experimental Results with a Case Study

A. A 45nm CMOS Clamped Bitline Sense Amplifier Circuit

The clamped bitline sense amplifier is a variation of the conventional sense amplifier used in DRAMs. The advantage of the clamped bitline is that it is clamped to a stable voltage after a sensing operation. This reduces the capacitive effect of the bitlines during the sensing operation, hence resulting in decreased dynamic power and sense delay [14], [15]. The schematic and physical design of the CBLSA are shown in Fig. 2(a) and Fig. 2(b). The initial design parameters for the transistors are length L_n, L_p = 45 nm, width W_n = 120 nm, and W_p = 240 nm. These dimensions are based on the nominal 45 nm technology node values and similar designs in [16]. The CBLSA needs matched transistors for optimal performance, making it a good test circuit to model the effects of process variation. The extracted SPICE netlist from the layout includes the parasitics of the design which impact its performance as seen in Table I.

B. Generation of Simple Kriging Metamodels for the FoMs

Each Kriging predicted point is calculated with a different weight. A parametric analysis using W_n and W_p as variables show that the circuit characteristics are dominated by W_n. The topology of the CBLSA circuit supports this trend as there are 10 NMOS transistors compared to 2 PMOS transistors. The operation of the circuit thus is more dependent on the variation of the NMOS transistor widths. The simple Kriging predicted surfaces are shown in Fig. 4.

Algorithm 2 Simulated-Annealing based optimization over simple Kriging metamodels of the CBLSA.

1: Initialize iteration counter: $counter \leftarrow 0$, Temperature: Θ and $Cooling\_Rate$, and start from a solution $\widehat{CBLSA}_i$.
2: Calculate FoMs for $\widehat{CBLSA}_i$ from Kriging metamodels.
3: Consider the objective T_{PC_i}. $result \leftarrow \Delta_{T_{PC}} \leftarrow T_{PC_i}$.
4: **while** ($\Delta_{T_{PC}} ! = 0$) **do**
5: $counter \leftarrow max\_Iteration$.
6: **while** ($counter > 0$) **do**
7: Make a random walk from $\widehat{CBLSA}_i$ to $\widehat{CBLSA}_j$.
8: Calculate FoMs for $\widehat{CBLSA}_j$ using the metamodels.
9: **if** ($T_{PCj} < result$) **then**
10: $result \leftarrow T_{PCj}$. $\widehat{CBLSA}_i \leftarrow \widehat{CBLSA}_j$.
11: **else**
12: $\Delta_{T_{PC}} \leftarrow T_{PC_i} - T_{PCj}$.
13: **if** ($\Delta_{T_{PC}} < 0$, random(0,1) $< e^{\frac{\Delta_{T_{PC}}}{T}}$) **then**
14: $T_{PCi} \leftarrow T_{PCj}$. $\widehat{CBLSA}_i \leftarrow \widehat{CBLSA}_j$.
15: **end if**
16: **end if**
17: $counter \leftarrow counter - 1$.
18: **end while**
19: $\Theta \leftarrow \Theta \times Cooling\_Rate$.
20: **end while**
21: **return** $result$ and $\widehat{CBLSA}_i$.

TABLE I
Characterization of the Baseline 45nm CBLSA Circuit Design.

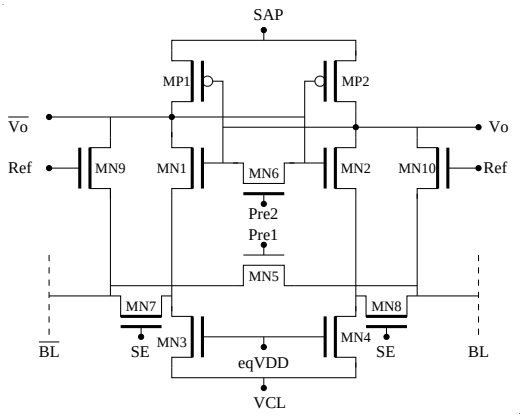
Design	Precharge Time T_{PC} (ns)	Sense Delay T_{SD} (ns)	Power P_{SA} (μW)	Sense Margin V_{SM} (mV)	Area (μm^2)
Schematic	10.31	1.79	1.84	26.91	-
Layout	10.40	1.91	1.88	26.86	6.045

C. Accuracy Analysis of Simple Kriging Metamodels

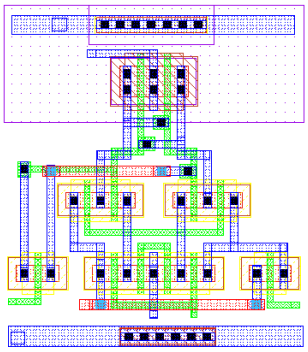
An exhaustive simulation was performed to compare the accuracy of the Kriging metamodels. A total of 1000 design points were simulated to densely capture a "golden surface" in the design space. A statistical analysis for the the metamodels shows that the accuracy is very high. A summary of the statistical analysis of the metamodels for 2-variable Kriging metamodels generated with 100 sample points is shown in Table II. The metrics used for analysis are the Root Mean Square Error (RMSE) and the correlation coefficient R^2. The correlation coefficient R^2 also known as the cross-correlation coefficient gives the quality of a least squares fitting compared to the original data. It essentially gives an estimate of the confidence level or how well the predicted metamodel will perform. A complete correlation gives R^2 of 1, so the closer to 1, R^2 is, the more accurate the metamodel.

TABLE II
STATISTICAL ANALYSIS OF THE SIMPLE KRIGING PREDICTED VALUES.

FoMs	Precharge Time, T_{PC}		Sense Delay, T_{SD}		Average Power, P_{SA}		Sense Margin, V_{SM}	
Samples	20	100	20	100	20	100	20	100
MSE	1.7×10^{-18}	7.6×10^{-20}	1.4×10^{-13}	5.6×10^{-06}	5.4×10^{-20}	1.4×10^{-22}	5.6×10^{-06}	7.9×10^{-08}
$RMSE$	1.3×10^{-09}	2.8×10^{-10}	3.8×10^{-07}	5.3×10^{-08}	2.3×10^{-10}	1.2×10^{-11}	2.4×10^{-03}	2.8×10^{-04}
R^2	0.9810	0.9838	0.9664	0.9673	0.9971	0.9986	0.9436	0.9548
STD	6.9×10^{-10}	6.9×10^{-10}	1.8×10^{-07}	4.5×10^{-08}	1.2×10^{-10}	9.3×10^{-12}	1.1×10^{-03}	2.1×10^{-04}



(a) Schematic design.



(b) Physical design.

Fig. 2. Circuit and layout for the clamped bitline sense amplifier.

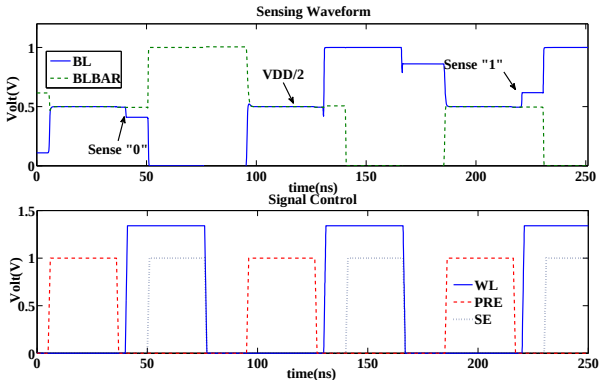
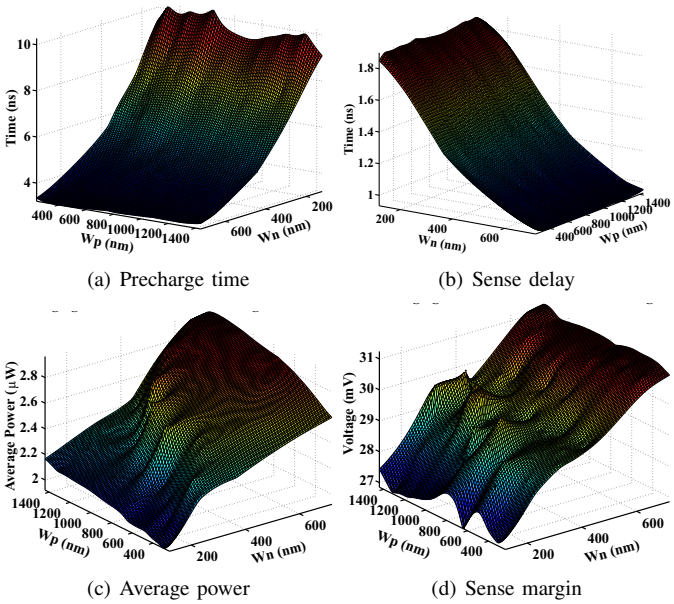


Fig. 3. Waveform of sense amplifier functional simulation.



(a) Precharge time (b) Sense delay

(c) Average power (d) Sense margin

Fig. 4. Simple Kriging predicted surfaces of the CBLSA using two variables (W_n and W_p) as the design parameters.

From an analysis of the results it is seen that the predicted points have an average R^2 of 0.97. The **simulation time for the generation of the simple Kriging metamodels was approximately 10.5 min. for the 2-variable metamodel compared to 72 hours for an exhaustive simulation**. The reduced time for sampling and creation of metamodels with very high accuracy demonstrates the efficiency of the simple Kriging metamodeling. The exhaustive simulation time increases exponentially as the number of variables increase, thus it is expected that the efficiency of this technique will be much better for higher dimensional and complex designs. A comparison of the simple Kriging and polynomial metamodels is shown in Table III. Simple Kriging metamodels consistently perform better than polynomial metamodels.

D. CBLSA Optimization Results

The temperature Θ is initially set to a high value, and a random walk is carried out at that temperature. Then the

TABLE III
COMPARISON OF KRIGING AND POLYNOMIAL METAMODELING.

FoMs	RMSE		R^2	
	Kriging	Polynomial	Kriging	Polynomial
T_{PC}	2.76×10^{-10}	5.07×10^{-10}	0.9838	0.9301
T_{SD}	5.27×10^{-08}	8.66×10^{-08}	0.9673	0.9046
P_{SA}	1.18×10^{-11}	5.91×10^{-11}	0.9986	0.9617
V_{SM}	2.81×10^{-04}	3.75×10^{-04}	0.9548	0.8925

temperature is lowered according to a cooling schedule. This initially high value of Θ allows the algorithm to accept new locations to search for optimal values even if they are worse than the current solution. As Θ declines, the probability of accepting worse solutions decreases but is not entirely zero. The slight probability of taking a step even if the new solution is worse is what allows simulated annealing to frequently get out of local minima. The algorithm stops when it reaches an acceptable solution or reaches the maximum number of iterations. The objective chosen for optimization is the precharge time T_{PC} with the average power consumption used as design constraint.

The simulated annealing algorithm is heuristic and does not give unique answers for each simulation. An average number of runs gives the finalized values in Tables IV. Each of the simulations were run for a maximum of 100 iterations, with each solution close in value but not unique. For the final CBLSA design, T_{PC} is reduced by 61.54% while P_{SA} was increased by 48.4%. For this design there is a larger improvement on the precharge time, however the average power is increased. T_{SD} and V_{SM} however are also improved significantly by 45.02% and 12.99%, respectively.

TABLE IV
FOMS OF THE OPTIMAL 45NM CBLSA DESIGN.

	Precharge Time T_{PC}	Sense Delay T_{SD}	Power P_{SA}	Sense Margin V_{SM}	Area
Optimal Design	4.04 ns	1.05 ns	2.79 μW	30.34 mV	6.356 μm^2
Change	61.15 %	45.02 %	-48.4 %	-12.99 %	5.15 %

The simulated annealing algorithm over the Kriging metamodels on average finds optimized values in 2.78 ms compared to a run of 45 minutes for an exhaustive search optimization on the metamodels. In other words, the proposed design flow could speedup the optimization process by several orders of magnitude. In general, with the generation of the metamodels, which includes the time for sampling the design, the design process takes approximately 11 min. compared to a 72-hour exhaustive simulation of the circuit without metamodels.

V. CONCLUSIONS AND FUTURE RESEARCH

A new methodology that uses simple Kriging metamodels and a simulated annealing algorithm for a sense amplifier optimization is presented. Simple Kriging metamodel based design optimization techniques provide the designer with methods to obtain accurate designs in a fast and efficient manner. They overcome the problem of local optimization because they provide robust models which are sufficiently accurate over the global design space. Experimental results on the metamodeling and design optimization of a CBLSA confirm the high accuracy of the generated metamodels with very low RMSE and high R^2. The optimization of CBLSA using the proposed flow improved its precharge time by 61.54% while speeding up the design process by 390$\times$. In future research, the methodology will be extended to high dimensional design parameters, incorporation of wire size along with the device size, and statistically modeling the effects of process variation. Optimization algorithms for multi-objective designs will also be explored.

REFERENCES

[1] O. Okobiah, S. P. Mohanty, E. Kougianos, and O. Garitselov, "Kriging-Assisted Ultra-Fast Simulated-Annealing Optimization of a Clamped Bitline Sense Amplifier," in *Proceedings of the 25th IEEE International Conference on VLSI Design (VLSID)*, 2012, pp. 145–150.

[2] B. Ankenman, B. Nelson, and J. Staum, "Stochastic Kriging for Simulation Metamodeling," in *Proceedings of the Winter Simulation Conference (WSC)*, Dec. 2008, pp. 362–370.

[3] C. Ling, L. Jia-hong, W. Bing, and H. Zhi-wei, "A Metamodel-Based Optimization Method for Mobile ad hoc Networks," in *Proceedings of International Conference on Computer Application and System Modeling (ICCASM)*, vol. 12, Oct. 2010, pp. V12–412 –V12–416.

[4] G. Yu and P. Li, "Yield-Aware Analog Integrated Circuit Optimization Using Geostatistics Motivated Performance Modeling," in *Proc. of International Conf. on Computer-Aided Design*, 2007, pp. 464–469.

[5] H. You, M. Yang, D. Wang, and X. Jia, "Kriging Model Combined with Latin Hypercube Sampling for Surrogate Modeling of Analog Integrated Circuit Performance," in *Proceedings of the International Symposium on Quality of Electronic Design*, 2009, pp. 554–558.

[6] N. A. C. Cressie, *Statistics for Spatial Data*. New York: Wiley, 1993.

[7] W. Van Beers, "Kriging Metamodeling in Discrete-Event Simulation: An Overview," in *Proceedings of the Winter Simulation Conference*, 2005, pp. 202–208.

[8] G. Matheron, "Principles of geostatistics", *Econ. Geol.*, vol. 58, no. 8, pp. 1246–1266, Dec. 1963.

[9] D. Ghai, S. P. Mohanty, and E. Kougianos, "Design of parasitic and process-variation aware nano-cmos rf circuits: A vco case study," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 17, no. 9, pp. 1339–1342, sept. 2009.

[10] K. T. Fang, R. Li, and A. Sudjianto, *Design and Modeling for Computer Experiments*. Boca Ration, FL: Chapman and Hall/CRC, 2006.

[11] O. Garitselov, S. P. Mohanty, and E. Kougianos, "A Comparative Study of Metamodels for Fast and Accurate Simulation of Nano-CMOS Circuits," *IEEE Transactions on Semiconductor Manufacturing*, vol. 25, no. 1, pp. 26–36, feb. 2012.

[12] *mGstat: A Geostatistical Matlab Toolbox*. [Online]. Available: mgstat.sourcefourge.net

[13] D. Gorissen, I. Couckuyt, P. Demeester, T. Dhaene, and K. Crombecq, "A Surrogate Modeling and Adaptive Sampling Toolbox for Computer Based Design," *J. Mach. Learn. Res.*, vol. 11, pp. 2051–2055, Aug 2010.

[14] Y. C. Lai and S. Y. Huang, "A Resilient and Power-Efficient Automatic-Power-Down Sense Amplifier for SRAM Design", *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 55, no. 10, pp. 1031–1035, Oct. 2008.

[15] T. N. Blalock and R. C. Jaeger, "A Subnanosecond Clamped-Bit-Line Sense amplifier for 1T Dynamic RAMs," in *Proc. of International Sympo. on VLSI Tech., Systems, & Applications*, 1991, pp. 82–86.

[16] O. Okobiah, S. P. Mohanty, E. Kougianos, and M. Poolakkaparambil, "Towards Robust Nano-CMOS Sense amplifier Design: A Dual-Threshold versus Dual-Oxide Perspective," in *Proceedings of the 21st ACM Great Lakes Symposium on VLSI*, 2011, pp. 145–150.