Citation	El-Moselhy, Tarek, and Luca Daniel. "Variation-aware Interconnect Extraction Using Statistical Moment Preserving Model Order Reduction." Design, Automation & Test in Europe Conference & Exhibition (DATE), 2010. 2010. 453-458. ©2010 IEEE.
As Published	
Publisher	Institute of Electrical and Electronics Engineers
Version	Final published version
Accessed	Sun Jan 21 21:29:39 EST 2018
Citable Link	http://hdl.handle.net/1721.1/62170
Terms of Use	Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.
Detailed Terms	

Variation-Aware Interconnect Extraction using Statistical Moment Preserving Model Order Reduction

Tarek El-Moselhy and Luca Daniel

Abstract—[1] **In this paper we present a stochastic model order reduction technique for interconnect extraction in the presence of process variabilities, i.e. variation-aware extraction. It is becoming increasingly evident that sampling based methods for variation-aware extraction are more efficient than more computationally complex techniques such as stochastic Galerkin method or the Neumann expansion. However, one of the remaining computational challenges of sampling based methods is how to simultaneously and efficiently solve the large number of linear systems corresponding to each different sample point. In this paper, we present a stochastic model reduction technique that exploits the similarity among the different solves to reduce the computational complexity of subsequent solves. We first suggest how to build a projection matrix such that the statistical moments and/or the coefficients of the projection of the stochastic vector on some orthogonal polynomials are preserved. We further introduce a proximity measure, which we use to determine *apriori* if a given system needs to be solved, or if it is instead properly represented using the currently available basis. Finally, in order to reduce the time required for the system assembly, we use the multivariate Hermite expansion to represent the system matrix. We verify our method by solving a variety of variation-aware capacitance extraction problems ranging from on-chip capacitance extraction in the presence of width and thickness variations, to off-chip capacitance extraction in the presence of surface roughness. We further solve very large scale problems that cannot be handled by any other state of the art technique.**

I. Introduction

Due to increasing process variations in modern fabrication technologies, it is becoming important to predict the effect of the geometrical variabilities on the electrical characteristics of the interconnect structures. This is typically achieved via the so-called variation-aware solvers [1], [2], [3]. The objective of such solvers is to efficiently compute the random unknown vector $\mathbf{x}(\mathbf{p})$ in a random linear system of equations $\mathbf{A}(\mathbf{p})\mathbf{x}(\mathbf{p}) = \mathbf{b}(\mathbf{p})$, where $\mathbf{p}$ is a large vector of random variables representing the process variations. There are two different categories of algorithms for achieving such objective, namely, "intrusive" and "non-intrusive" algorithms.

The term "intrusive" refers to the fact that the unknown vector $\mathbf{x}(\mathbf{p})$ is computed using specialized stochastic solvers. Examples of intrusive algorithms are the Neumann expansion [1], the stochastic finite element simulation (SFE) [4], and the combined Neumann-Hermite expansion [3].

On the other hand, "non-intrusive" algorithms are those that rely on sampling the parameter space vector $\mathbf{p}$, and then solving a deterministic system $\mathbf{A}(\mathbf{p}_k)\mathbf{x}(\mathbf{p}_k) = \mathbf{b}(\mathbf{p}_k)$ for every sample $\mathbf{p}_k$ of the parameter space. The term "non-intrusive" refers to the fact that any standard deterministic solver can be used to compute the unknown vector $\mathbf{x}(\mathbf{p}_k)$ in each solve. Examples of non-intrusive algorithms include the well-known Monte Carlo method, and the stochastic collocation method [5].

One single solve of an intrusive algorithm is typically much more expensive than one single solve of a corresponding deterministic solver. It is indeed true that one intrusive solve can typically produce all the desired statistics of the unknown. However it is also true that the many deterministic solves required by non-intrusive methods could require less time than one intrusive solve, provided the following two issues are addressed effectively: the optimal selection of sample points in a large dimensional parameter space, and *efficient* solution of the "similar" linear systems $\mathbf{A}(\mathbf{p}_k)\mathbf{x}(\mathbf{p}_k) = \mathbf{b}(\mathbf{p}_k)$. Acceptable solutions to the first challenge vary from sparse grid sampling [5] to Monte Carlo sampling. The state of the art with regard to the second question is to use Krylov subspace recycling [6].

In this paper we claim the field of parameterized linear model order reduction (PROM) can be related to the problem of variation-aware extraction. The problem description is given via a parameterized dynamical linear system. However, if we remove the time dependence, the setup for the variation-aware extraction problem and the parameterized model reduction problem becomes identical. The problem that is typically addressed in PROM research is how to choose a basis that spans the solution space. There are two different ways of constructing such basis. The first is to use deterministic moment matching [7], which has proven to be very inefficient in large dimensional parameter spaces due to the curse of dimensionality. The alternative, more efficient approach is to use multi-point matching. In multi-point matching the basis is constructed from the solution of the linear system at different sampling points in the parameter space. A lot of research has been dedicated to developing efficient algorithms for sampling high dimensional parameter spaces, such that the constructed basis appropriately span the solution domain [8], [9], [10]. The question of which properties of the original system are being preserved in the reduced system, when using multi-point projection methods, has been barely addressed.

As obvious from the discussion above there is a clear similarity between non-intrusive variation-aware solvers and multi-point model reduction methods, since both of these techniques inherently rely on solving the given linear system at different points in the parameter space. In this paper we develop a multi-point matching approach based on quadrature point sampling that guarantees the preservation of the statistical moments of the quantity of interest, or equivalently the coefficients of the projection of the statistical quantities on some space of orthogonal polynomials. Furthermore, we make the connection between parameterized model reduction and non-intrusive sampling methods for variation-aware extraction. Finally, we make the connection and point out the differences between variational-TBR [10] and our proposed method for variation-aware extraction.

In section III we demonstrate that important statistical characteristics of the solution can be preserved by using a specially constructed projection matrix. In section IV we discuss several techniques to reduce the computational complexity of our approach. Finally, in section VI we demonstrate the validity of our approach on a variety of examples, especially on a large size example that cannot be solved using any of the other state of the art approaches.

II. BACKGROUND

A. Electromagnetic Formulations for Capacitance Extraction

In this section we will focus primarily on the electro-quasi-static capacitance extraction using the first type Fredholm integral equation, however the techniques presented in this paper can be equally applied to any general full impedance extraction formulation, such as the mixed potential integral equation. Given a set of conductors, the relation between the given conductor surface electric potential $\phi(\mathbf{r})$, and the unknown surface charge distribution $\rho(\mathbf{r})$ is

$$\frac{1}{4\pi\varepsilon} \int_S G(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}')\mathbf{dr}' = \phi(\mathbf{r}) \qquad (1)$$

where $G(\mathbf{r},\mathbf{r}')$ is the Green's function, S is the conductors surface area, and ε is the complex dielectric constant including dielectric losses.

A standard procedure for solving (1) involves discretizing the unknown charge density $\rho(\mathbf{r})$ using piecewise constant basis functions and Galerkin testing to construct the linear system

$$\mathbf{A}(\mathbf{p})\,\mathbf{x}(\mathbf{p}) = \mathbf{b} \qquad (2)$$

where $\mathbf{A}$, $\mathbf{x}$ and $\mathbf{b}$ are described in detail in [11]. We assume that the N_P parameters in vector $\mathbf{p}$ describe variations in geometry or material properties and can be treated as correlated random variables. For "off-chip" interconnects, $\mathbf{p}$ might describe for instance the roughness of the conductors' surfaces S, which we assume described by a stochastic Gaussian process, and by a Gaussian correlation function:

$$\mathcal{P}(\mathbf{p}) = \frac{\exp(-0.5\mathbf{p}^T\Sigma^{-1}\mathbf{p})}{(2\pi)^{0.5n}\sqrt{|\Sigma|}}, \quad \Sigma_{ij} = \sigma^2 \exp\left(-\frac{\|\mathbf{r}_i-\mathbf{r}_j\|^2}{L_c^2}\right),$$

where L_c is the correlation length, $p_k \in \mathbf{p}$ is the surface height at location $\mathbf{r}_k$, and $\Sigma \in \mathbb{R}^{N_p \times N_p}$ is the correlation matrix.

For the case of "on-chip" interconnect variation, the random parameters in $\mathbf{p}$ represent uncorrelated Gaussian variations in the geometrical dimensions (e.g. width and height) of the interconnects. Consequently, the joint probability density function of such random Gaussian variations is a product of the individual random variations.

In order to simplify subsequent analysis, the set of geometrical random variables, $\mathbf{p}$, is transformed into a set of uncorrelated Gaussian random variables $\vec{\eta}$ using the Karhunen Loeve expansion [12]. Consequently, equation (2) is transformed to:

$$\mathbf{A}(\vec{\eta})\,\mathbf{x}(\vec{\eta}) = \mathbf{b} \qquad (3)$$

Typically the quantity of interest (e.g. the capacitance) can be then expressed as a linear combination of the state vectors

$$y(\eta) = \mathbf{c}^T\mathbf{x}(\eta) \qquad (4)$$

B. Non-Intrusive Sampling Methods

In non-intrusive sampling based methods, the system (3) is solved N_S times, for different realizations of the random process $\vec{\eta}$, obtaining the set of solutions $\{\mathbf{x}(\vec{\eta}_1),\ \mathbf{x}(\vec{\eta}_2),\ \cdots,\ \mathbf{x}(\vec{\eta}_{N_S})\}$. The statistical characteristics of $\mathbf{x}(\vec{\eta})$ are then computed from the obtained solution set. The computational inefficiency of sampling based methods stems from the need for a very large number N_S of sample points, and consequently system solves, to cover the entire parameter space. For instance, a very large number of samples is needed in Monte-Carlo sampling, since the convergence is very slow $O(\frac{1}{\sqrt{N_S}})$, and in sparse grid sampling, since the number of samples grows polynomially with the dimension of the parameter space.

One fundamental characteristic that we exploit in the rest of this paper is that the linear systems produced by the sampling methods are all "similar". This similarity is due to the fact that all underlying systems share the same geometry and have comparable parameter values.

C. Projection-based Model Reduction

In projection methods the unknown is assumed to live in a subset of the N-dimensional space spanned by a small number of basis vectors, i.e. $\mathbf{x}_r = \mathbf{Uz}$, where the columns of $\mathbf{U} \in \mathbb{R}^{N \times r}$ represent a collection of r basis vectors. Using $\mathbf{x}_r$ and some left projection matrix $\mathbf{V} \in \mathbb{R}^{r \times N}$, the reduced system and output become:

$$\begin{aligned} \mathbf{VAUz} &= \mathbf{Vb} \\ y_r &= \mathbf{c}^T\mathbf{U}(\mathbf{VAU})^{-1}\mathbf{Vb} \end{aligned}$$

One standard way to construct the projection matrix $\mathbf{U}$ is to use multi-point matching techniques in which the columns of $\mathbf{U}$ are the solutions of the linear system at different values of the parameter vector. Note that the similarity between projection based model reduction methods and non-intrusive stochastic simulation methods stems from the fact that both methods rely on solving the linear system at multiple points in the parameter space.

One standard way to construct the left projection matrix is to let $\mathbf{V} = \mathbf{U}^H$, i.e. use a Galerkin projection

$$y_r = \mathbf{c}^T \mathbf{U} \left(\mathbf{U}^H \mathbf{A} \mathbf{U} \right)^{-1} \mathbf{U}^H \mathbf{b} \qquad (5)$$

III. STOCHASTIC MODEL REDUCTION

The question remains on how to determine the set of sample points which is used to construct the basis. The objective is to construct the basis such that important properties of the original system are preserved in the reduced system. In stochastic problems, such as variation-aware extraction, we propose to preserve *statistical* moments:

$$
\begin{aligned}
E\left[y^k\right] &= \int y^k w(\vec{\eta}) d\vec{\eta} = \int \left(\mathbf{c}^T \mathbf{A}(\vec{\eta})^{-1} \mathbf{b}\right)^k w(\vec{\eta}) d\vec{\eta} \\
&\simeq \sum_{q=1}^{N_q} \alpha_q (\mathbf{c}^T \mathbf{A}(\vec{\eta}_q)^{-1} \mathbf{b})^k w(\vec{\eta}_q) \qquad (6)
\end{aligned}
$$

where $w(\vec{\eta})$ is the multivariate probability density function (PDF) associated with $\vec{\eta}$, α_q are the weights associated with the quadrature points $\vec{\eta}_q$ and N_q is the total number of quadrature points required by the numerical quadrature scheme used to calculate the integral (N_q increases as k increases). In general we will be interested in moments up to second order (i.e. $k = 2$), which means we will be able to use relatively low order quadrature schemes.

Instead of (or in addition to) preserving statistical moments, based on the application, we might be also interested in matching the coefficients β_i of the projection of the solution $y(\vec{\eta}) = \sum_{i=1}^{K} \beta_i h_i(\vec{\eta})$ on the space of multivariate polynomials, $h_i(\vec{\eta})$, orthogonal with respect to the probability measure $w(\vec{\eta})$, e.g. Hermite polynomials [13]:

$$
\begin{aligned}
\beta_i = \langle y, h_i(\vec{\eta}) \rangle &= \int y h_i(\vec{\eta}) w(\vec{\eta}) d\vec{\eta} \\
&\simeq \sum_{q=1}^{N_q} \alpha_q \mathbf{c}^T \mathbf{A}(\vec{\eta}_q)^{-1} \mathbf{b} h_i(\vec{\eta}_q) w(\vec{\eta}_q) \quad (7)
\end{aligned}
$$

Theorem III.1 below provides a way to achieve both objectives, i.e. matching of the statistical moments, and matching of the coefficients for a multivariate Hermite representation of the solution.

Theorem III.1. *IF* $\mathbf{U}$ *is an orthonormal projection matrix and*

$$colspan\{\mathbf{U}\} \supset span\left\{\mathbf{b}, \mathbf{A}(\vec{\eta}_1)^{-1}\mathbf{b}, \mathbf{A}(\vec{\eta}_2)^{-1}\mathbf{b}, \cdots, \mathbf{A}(\vec{\eta}_{N_q})^{-1}\mathbf{b}\right\}$$

where $\left\{\vec{\eta}_q : q = 1, \cdots, N_q\right\}$ *is a set of quadrature points such that (6) and (7) are exactly integrable for some choice of orthogonal polynomials* $h_i(\vec{\eta})$ *THEN*

$$E\left[y^k\right] = E\left[y_r^k\right] \qquad (8)$$

and

$$\langle y, h_i \rangle = \langle y_r, h_i \rangle \qquad (9)$$

Proof: Using a quadrature rule with the given points on the reduced system we obtain:

$$
\begin{aligned}
E\left[y_r^k\right] &= \sum_{q=1}^{N_q} \alpha_q \left(\mathbf{c}^T \mathbf{U}(\mathbf{U}^H \mathbf{A}(\vec{\eta}_q)\mathbf{U})^{-1}\mathbf{U}^H \mathbf{b}\right)^k w(\vec{\eta}_q) \\
\langle y_r, h_i(\vec{\eta}) \rangle &= \sum_{q=1}^{N_q} \alpha_q \mathbf{c}^T \mathbf{U}(\mathbf{U}^H \mathbf{A}(\vec{\eta}_q)\mathbf{U})^{-1}\mathbf{U}^H \mathbf{b} h_i(\vec{\eta}_q) w(\vec{\eta}_q)
\end{aligned}
$$

Using first lemma 3, and then lemma 2 in [7] the common term in both expressions becomes:

$$\mathbf{c}^T \mathbf{U}(\mathbf{U}^H \mathbf{A}(\vec{\eta}_q)\mathbf{U})^{-1}\mathbf{U}^H \mathbf{b} = \mathbf{c}^T \mathbf{U}\mathbf{U}^H \mathbf{A}(\vec{\eta}_q)^{-1}\mathbf{b} = \mathbf{c}^T \mathbf{A}(\vec{\eta}_q)^{-1}\mathbf{b}$$

therefore

$$
\begin{aligned}
E\left[y_r^k\right] &= \sum_{q=1}^{N_q} \alpha_q (\mathbf{c}^T \mathbf{A}(\vec{\eta}_q)^{-1}\mathbf{b})^k w(\vec{\eta}_q) = E\left[y^k\right] \\
\langle y_r, h_i(\vec{\eta}) \rangle &= \sum_{q=1}^{N_q} \alpha_q \mathbf{c}^T \mathbf{A}(\vec{\eta}_q)^{-1}\mathbf{b} h_i(\vec{\eta}_q) w(\vec{\eta}_q) = \langle y, h_i(\vec{\eta}) \rangle
\end{aligned}
$$

∎

Theorem III.1 does not rely on a particular integration rule. Consequently we are at liberty to choose any of the available numerical integrations schemes, including but not limited to Monte Carlo, Quasi Monte Carlo, importance sampling, tensor products, or sparse grid products. Furthermore, since what we really want is to accurately represent the space spanned by $\mathbf{U}$, rather than to accurately compute the integral, we are at liberty to use even those complicated optimization methods for sampling point selections [8]. For simplicity in this paper we will construct $\mathbf{U}$ using sparse grid quadrature schemes if the parameter space dimension is less than 100 and Monte Carlo integration otherwise.

IV. PROXIMITY MEASURE FOR COMPUTATIONAL EFFICIENCY

Stacking the solutions at all quadrature points in one projection matrix would result in a large dimensional ill-conditioned projection matrix. Furthermore, obtaining in the first place the solutions corresponding to all quadrature points requires solving many large linear systems. The standard solution for the first two problems (which does not address the third one) would be to compress the projection subspace using a singular value decomposition (SVD).

We propose here instead an alternative solution that addresses *all* three problems at the same time. Specifically, before solving a particular large system at a new quadrature point, we propose to compute instead the following residual

$$\mathbf{r}(\vec{\eta}) = \mathbf{b} - \mathbf{A}(\vec{\eta})\mathbf{x}_r(\vec{\eta}) = \mathbf{b} - \mathbf{A}(\vec{\eta})\mathbf{U} \left(\mathbf{U}^H \mathbf{A}(\vec{\eta})\mathbf{U}\right)^{-1}\mathbf{U}^H \mathbf{b}$$

If the norm of the residual is large, then the current subspace does not faithfully represent the solution and therefore needs to be expanded with the solution of the new system orthogonalized with respect to the current subspace. If, on the other hand, the norm is small then the solution at the new sampling point is accurately represented using the currently explored subspace and does not need to be added to the basis. The

advantage of using such a proximity measure is that only *one* matrix-vector product is done in the original space $O(N^2)$ and all the rest of the computation, i.e. computing $\mathbf{x}_r(\vec{\eta})$, is done in the reduced space, and is therefore very efficient. Notice that a similar residual measure was used in [9] to determine the points at which the parameter space should be sampled. Algorithm 1 summarizes our complete stochastic model reduction variation-aware extraction approach.

Algorithm 1 Stochastic Model Reduction Method for Solving Linear Systems with Random Matrices.

1: $\mathbf{U} \leftarrow \mathbf{b}$
2: $q \leftarrow 0$
3: **for** each quadrature point $\vec{\eta}_q$ **do**
4:     $q \leftarrow q + 1$
5:     generate linear system $\mathbf{A}(\vec{\eta}_q)$.
6:     compute $\mathbf{x}_r(\vec{\eta}_q) \leftarrow \mathbf{U}\left(\mathbf{U}^H \mathbf{A}(\vec{\eta}_q)\mathbf{U}\right)^{-1}\mathbf{U}^H\mathbf{b}$
7:     compute residual $\mathbf{r}(\vec{\eta}_q) \leftarrow \mathbf{b} - \mathbf{A}(\vec{\eta}_q)\mathbf{x}_r(\vec{\eta}_q)$.
8:     **if** $\frac{\|\mathbf{r}(\vec{\eta}_q)\|}{\|\mathbf{b}\|} > $ theshold **then**
9:         solve for $\mathbf{x}(\vec{\eta}_q)$, $\mathbf{A}(\vec{\eta}_q)\mathbf{x}(\vec{\eta}_q) = \mathbf{b}$
10:        extend the basis $\mathbf{U}$ with $\mathbf{x}(\vec{\eta}_q) - \mathbf{U}\mathbf{U}^H\mathbf{x}(\vec{\eta}_q)$
11:    **end if**
12: **end for**

The complexity of computing the Galerkin projection can be reduced by first using the multivariate Hermite expansion to express the system matrix in terms of orthogonal polynomials

$$\mathbf{A}(\mathbf{p}) = \sum_{i=1}^{K} \mathbf{A}_i h_i(\vec{\eta}) \tag{10}$$

The coefficients of the expansion $\mathbf{A}_i$ are the projection of the matrix on the space of orthogonal polynomials.

$$\mathbf{A}_i = \langle \mathbf{A}(\vec{\eta}), h_i(\vec{\eta}) \rangle = \int \mathbf{A}(\vec{\eta}) h_i(\vec{\eta}) w(\vec{\eta}) d\vec{\eta}$$

An efficient technique to compute the multidimensional integral has been provided in [3].

Using the multivariate Hermite expansion (10), we observe that each time the basis $\mathbf{U}_{r+1} = \begin{bmatrix} \mathbf{U}_r & \mathbf{u} \end{bmatrix}$ is expanded from $\mathbf{U}_r$ by $\mathbf{u}$, the reduced system matrix can be updated as follows

$$\mathbf{U}_{r+1}^H \mathbf{A}(\vec{\eta})\mathbf{U}_{r+1} = \sum_{i=1}^{K} \mathbf{U}_{r+1}^H \mathbf{A}_i \mathbf{U}_{r+1} h_i(\vec{\eta})$$

where

$$\begin{bmatrix} \mathbf{U}_r^H \\ \mathbf{u}^H \end{bmatrix} \mathbf{A}_i \begin{bmatrix} \mathbf{U}_r & \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{U}_r^H \mathbf{A}_i \mathbf{U}_r & \mathbf{U}_r^H \mathbf{A}_i \mathbf{u} \\ \mathbf{u}^H \mathbf{A}_i \mathbf{U}_r & \mathbf{u}^H \mathbf{A}_i \mathbf{u} \end{bmatrix} \tag{11}$$

Since the terms $\mathbf{U}_r^H \mathbf{A}_i$, $\mathbf{A}_i \mathbf{U}_r$ and $\mathbf{U}_r^H \mathbf{A}_i \mathbf{U}_r$ are readily available, we only need to compute $\mathbf{A}_i \mathbf{u}$ and $\mathbf{u}^H \mathbf{A}_i$.

A. Computational Complexity

The time complexity of generating the reduced model is $N_S O(N^2)$, which means that the asymptotic complexity of using model reduction is equal to that of solving a total of N_S linear systems. However, the constant in front of the complexity is significantly reduced since in our approach

a *single* matrix product is required to validate the model (compute $\mathbf{A}\mathbf{x}_r$), whereas in the standard method a *complete* system solve has to be performed.

One of the main computational advantages of our method is that in large dimensional problems an accurate model is constructed using a number of solves much smaller than that required by non-intrusive stochastic collocation method. The unknowns at many of the sampling points can therefore be computed using the reduced model, which is computationally extremely efficient. Consequently, the factor N_S in the complexity of our method is typically much smaller than that in the standard stochastic collocation method.

V. RELATION TO NON-INTRUSIVE METHODS AND TO VARIATIONAL TBR

The main connection between our method, the non-intrusive stochastic collocation method (SCM) [5], and the variational TBR (vTBR) [10] is the fact that the sample points are generated based on a suitable quadrature scheme. In both our method and in SCM, any number of statistical moments can be explicitly preserved. This allows for proper choice of the integration scheme based on the order of the statistical moments to be preserved. On the other hand, in vTBR first and second order statistics are coincidentally enforced, since vTBR approximates the Grammian (which is also the covariance matrix) using quadrature. Consequently, any statistical moments of more than second order cannot be preserved. Furthermore, if the objective is to preserve just first order statistics then vTBR results in extra unnecessary computations since it implicitly preserves second order statistics. Consequently, variational-TBR can be considered a special case of our more general statistical moment matching technique in which the covariance matrix is preserved $E\left[\mathbf{x}^T\mathbf{x}\right]$.

On the other hand, both our method and vTBR use explicitly the projection approach. This allows for the efficient truncation of the model by finding a reduced basis to represent the column span of the solution vectors at the different sample points. This is not the case of the stochastic collocation method, in which the solution at *all* the sample points are explicitly used to construct the model. Consequently, for large dimensional parameter spaces the stochastic collocation method becomes inefficient.

VI. RESULTS

It is important to notice that problems encountered in the parameterized model reduction literature typically address problems with a small number of parameters (e.g. < 20) and their objective is to end up with models of order small order (e.g. $O(10)$). On the other hand, in variation-aware extraction we are instead faced with problems with a very large number of parameters (> 100), in which ending up with reduced models of order 1000 can still substantially speed up the overall computation, since such models will be computationally significantly faster to simulate than any available non-intrusive or intrusive method. In addition, while for parameterized model order reduction the model generation cost is less of an issue because it can be done offline,
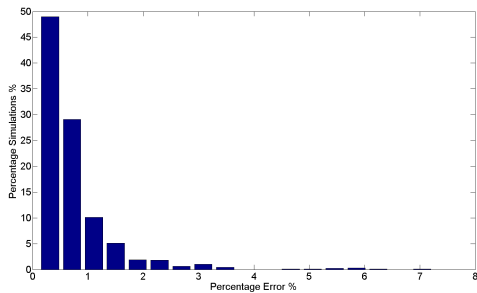
Fig. 1. Histogram of the error of the proposed model at 1000 randomly generated points in the parameter space for the small grid example

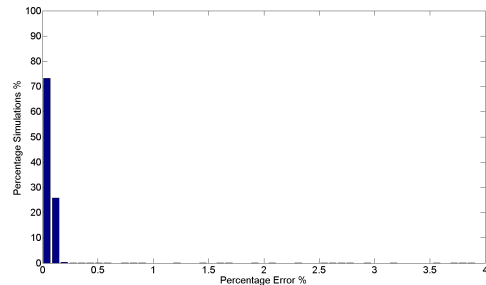Fig. 2. Histogram of the error of the proposed model at 7000 randomly generated points in the parameter space for the large grid example

in variation-aware extraction the model generation cost is absolutely critical since it typically represents more than 90% of the total computation time.

A. On-Chip Parameter Variation

In this section we show a typical on-chip parasitic extraction example with 16 conductors. In a cross-section view the conductors are arranged on a 4×4 regular grid. This geometry is often used to construct the capacitance look-up table necessary for efficient full chip extraction. The nominal dimensions of the conductors are 42nm width and 31nm height. The nominal separation between the conductors is 42nm. Each conductor is discretized in 30 segments, resulting in a total of N=481 unknowns (including one unknown potential since this is a 2D problem). The width of each conductor is assumed an independent Gaussian random variable with standard deviation 15.75nm. We assume that a second order stochastic expansion can accurately represent the variations in both the system matrix and the vector $\mathbf{x}$. We use a total number of 561 quadrature points to construct the basis. Such quadrature points are placed according to a second order sparse grid Smolyak construction. When using our proximity measure to avoid solving systems that do not add sufficiently to the column span of the projection matrix, our approach only generated a total of only 32 basis.

The total time required by our appraoch is 198 seconds as opposed to 436 seconds required by the standard SCM, and 52 seconds required the fast stochastic Galerkin method (SGM) [14]. Consequently, for this particularly small example our approach is 2.2× faster than the SCM, and 3.8× slower than the SGM.

We tested the accuracy of our reduced model by computing the solutions at 1000 randomly generated points, and comparing them to the full system. Figure 1 shows a histogram of the resulting error. We observe that the error is less than 3% for 98% of the simulated structures. We then used our reduced model to compute the second order statistical moments and all second order Hermite coefficients of the projection of the output capacitances, and observed a 1% match compared to the SCM and to the accelerated SGM.

B. Large On-Chip Example

This is an example of 100 conductors arranged on a 10×10 grid. Each conductor is of size 45nm $\times$ 30nm. Each conductor

is discretized using a total of 50 unknowns resulting in a total of 5000 unknowns. The width of each conductor is assumed to be an independent Gaussian random variable with variance of 20% the nominal width. The total number of random variables is 100. Using a second order sparse grid construction would require more than 20,000 simulations. Instead we choose to use a total of 5000 Monte Carlo simulations to build the basis and another 5000 simulations to test the accuracy of the model. We observe that only 112 basis are required to achieve an accuracy of better than 2% in the residual estimation for 99.5% of the cases under study (see figure 2). This means that our algorithm is about 130 times faster that the sparse grid technique and at least 65 times faster than Monte Carlo method. On the other hand, the SGM cannot be applied to such a large example due to memory requirements (more than 1TB). The statistical moments up to first order and all coefficients of the projection of the output capacitances on the space of first order Hermite polynomials, computed using our approach and the standard collocation method match up to 1%. Accuracy of higher order statistics could not be validated, since it is not possible to run a stochastic collocation method on this problem to be used as a reference.

C. Off-Chip Surface Roughness

This example is a parallel place capacitor, where the upper surface is characterized by a rough surface (see figure 3). The rough surface is described by a Gaussian PDF of $\sigma = 1$, and correlation length $L = 1$. The plate surfaces are $20L \times 20L$. The distance between the plates is $5L$. The total number of unknowns discretizing both plates is 21,000. The total number of independent random variables describing the stochastic surface is 323, which is orders of magnitude larger than anything available in the literature. In particular, the maximum number of parameters in [8], [9], [10] is 21, 4, 6 respectively. The total number of orthogonal polynomials required for a second order description of the stochastic capacitance is 52,650. The total number of sparse grid nodes, i.e. independent solves, required for a second order accurate expansion is more than 200,000. This number is even larger than $N = 21,000$. We therefore opted to use a Monte Carlo sampling method. We used 1,500 samples, and our algorithm produced a projection matrix that is of size 997. This projection matrix is then used to compute the solution of a different set of 1,500 instantiations of the

Example	Method	Time	Memory	Accuracy
	SCM	436	1.85MB	second order
Ex. A	SGM	52	11MB	second order
	SROM	198	1.85MB	higher order
	SCM	(100 hour)	4GB	second order
Ex. C	SGM	N.A.	165TB	second order
	SROM	1 hour	4GB	higher order

Fig. 3.   Large off-chip parallel plate capacitor with rough surface

Fig. 4.   Histogram of the error of the proposed model at 7000 randomly generated points in the parameter space

Fig. 5.   Probability density function of mutual capacitance, mean capacitance is approximately 40% larger than that of the parallel plate capacitor without rough surface
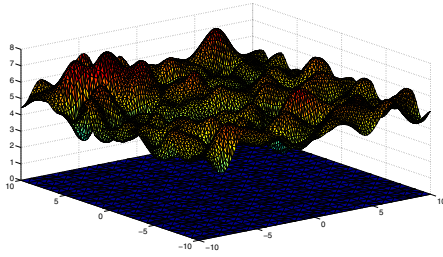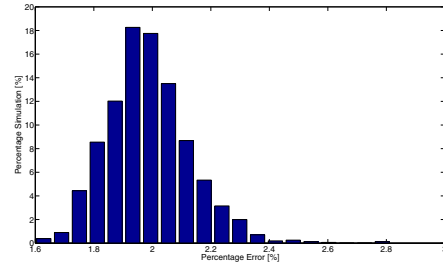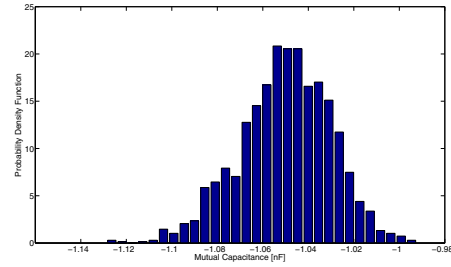
structure. Each instantiation is solved using fastcap [11]. The results computed using the reduced model and using the exact model are compared in figure 4. We observe that 99.9% of the cases exhibit error less than 5%. The probability density function of the mutual capacitance between both plates is shown in figure 5. The non-Gaussian nature of the curve is clearly visible and the deviation between maximum and minimum capacitances is about 10% from the mean. Table I shows that our algorithm is two orders of magnitude more efficient than the stochastic collocation method, and that the fast SGM cannot be applied for this algorithm due to its memory requirements.

VII. CONCLUSION

In this paper we have developed a new stochastic model order approach to increase the computational efficiency of the non-intrusive sampling based methods, such as stochastic collocation method, and the Monte Carlo method. We have developed a way to generate the projection matrix such that the statistical moments are preserved. Furthermore, we have suggested a proximity measure that can be used to identify "similar" systems. We have utilized the Hermite expansion to reduce the complexity of computing the reduced system from the original one. Finally, we have demonstrated the validity of our approach on a variety of problems in particular we have solved a very large problem with $N = 21,000$ unknowns and $N_P = 323$ parameters. The algorithm proposed in this paper is currently, to the best of our knowledge, the only available method that can actually solve such large scale problems.

REFERENCES

[1] Z. Zhu and J. White, "FastSies: A Fast Stochastic Integral Equation Solver for Modeling the Rough Surface Effect" *International Conference on Computer Aided Design 2005*.
[2] H. Zhu, X. Zeng, W. Cai, J. Xue and D. Zhou, "A sparse grid based spectral stochastic collocation method for variations-aware capacitance extraction of interconnect under nanometer process technology", *Design Automation and Test in Europe, 2007*.
[3] T. El-Moselhy and L. Daniel "Stochastic Integral Equation Solver for Efficient Variation-Aware Interconnect Extraction" *ACM/IEEE Design Automation Conference*, 2008.
[4] R. Ghanem and P. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Spring-Verlag, 1991.
[5] D. Xiu and J. Hesthaven, "High-Order Collocation Method for Differential Equations with Random Inputs" *SIAM J. Sci. Comput.*, 2005.
[6] M. Parks, E. De Sturler, G. Mackey, D. Johnson and S. Maiti " Recycling Krylov Subspaces for Sequences of Linear Systems" *SIAM Journal on Scientific Computing*, Vol. 28, 2006.
[7] L. Daniel O. C. Siong, L. S. Chay, K. H. Lee and J. White, "A Multiparameter Moment Matching Model Reduction Approach for Generating Geometrically Parameterized Interconnect Performance Models" *IEEE Trans. Computer-Aided Design*, 23:678693, May 2004.
[8] T. Bui-Thanh, K. Willcox, and O. Ghattas, "Model Reduction for Large-Scale Systems with High-Dimensional Parametric Input Space", *SIAM Journal on Scientific Computing*, Vol. 30, No. 6, pp. 3270-3288. 2008.
[9] J. F. Villena and L. M. Silveira, "ARMS - Automatic Residue-Minimization based Sampling for Multi-Point Modeling Techniques" *Design Automation conference*, 2009
[10] J. Phillips, "Variational interconnect analysis via PMTBR" *International Conference on Computer Aided Design*, 2004.
[11] K. Nabors and J. White, "FASTCAP A multipole-accelerated 3-D capacitance extraction program", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 10, pp. 1447-1459, November 1991.
[12] M.Loeve, *Probability Theory*, Spring-Verlag, 1977.
[13] D. Xiu and G. Karniadakis, "The Wiener-Askey Polynomial Chaos Expansion For Stochastic Differential Equations", *SIAM Journal of Scientific Computing*, October 2002.
[14] T. El-Moselhy and L. Daniel, "Efficient Solutions of Stochastic Linear Systems of Equations for Variation-Aware Impedance Extraction", *submitted to IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*