Statistical Analysis of On-Chip Power Grid Networks by Variational Extended Truncated Balanced Realization Method *

Duo Li†, Sheldon X.-D. Tan†, Gengsheng Chen$^\delta$ and Xuan Zeng$^\delta$
†Department of Electrical Engineering, University of California, Riverside, CA 92521
$^\delta$ASIC & System State-Key-Lab, Microelectronics Dept., Fudan Univeristy, Shanghai, China, 200433

ABSTRACT

In this paper, we present a novel statistical analysis approach for large power grid network analysis under process variations. The new algorithm is very efficient and scalable for huge networks with a large number of variational variables. This approach, called *varETBR* for variational extended truncated balanced realization, is based on model order reduction techniques to reduce the circuit matrices before the variational simulation. It performs the parameterized reduction on the original system using variation-bearing subspaces. *varETBR* calculates variational response Gramians by Monte-Carlo based numerical integration considering both system and input source variations for generating the projection subspace. *varETBR* is very scalable for the number of variables and is flexible for different variational distributions and ranges as demonstrated in experimental results. After the reduction, Monte-Carlo based statistical simulation is performed on the reduced system and the statistical responses of the original system are obtained thereafter. Experimental results, on a number of IBM benchmark circuits [15] up to 1.6 million nodes, show that the *varETBR* can be 4500*X* faster than the Monte-Carlo method and is much more scalable than one of the recently proposed approaches.

1. INTRODUCTION

Reliable on-chip power delivery is one of the major concerns for 90nm and below VLSI technology. This situation becomes worse as technology continues to scale to 45nm and below owning to the increasing process-induced variability [22]. The process induced variations manifest themselves at different levels (wafer level, die-level and within a die) and they are caused by different sources (lithograph, materials, aging etc) [3, 14]. Some of the variations are systematic like those caused by chemical mechanical polishing (CMP). Some are purely random like the doping density of impurities and edge roughness. As the technology moves to 65nm and comes near to 45nm, the variation will become more and more pronounced for both systemic and random components.

A number of research works have been proposed recently to address the voltage drop issues in the on-chip power delivery networks under process variations. The voltage drop of power grid networks subject to the leakage current variations were first studied in [5, 6]. This method assumes that the log-normal distribution of the node voltage drop is caused by log-normal leakage current inputs and is based on a localized Monte-Carlo (sampling) method to compute the variance of the node voltage drop. However, this localized sampling method is limited to the static DC solution of power grids modeled as resistor-only networks. Therefore, it can only compute the responses to the standby leakage currents. However, the dynamic leakage currents become more significant, especially when the sleep transistors are intensively used nowadays for reducing leakage powers. In [23, 17], impulse responses are used to compute the means and variances of node voltage responses caused by general current variations. But this method needs to know the impulse response from all the current sources to all the nodes, which is expensive to compute for a large network. Methods proposed in [8, 7] using orthogonal polynomial chaos expansion of random processes to represent and solve for the stochastic responses of linear systems. But existing approaches can only consider the Gaussian distributions and analysis times increases with the number of variables. The methods have been improved by the StoEKS method [12], where reduction is performed on the variational circuit matrices before the simulation.

In this paper, we present a novel scalable statistical simulation approach for large power grid network analysis considering process variations. The new algorithm is very scalable for large networks with a large number of random variables. The new method, called *varETBR*, is based on the recently proposed extended truncated balanced realization (ETBR) method [10]. To consider the variational parameters, Monte-Carlo like numerical integration is applied, which is similar to the varPMTBR method [18]. But different from varPMTBR, varETBR calculates the variational response Grammians, considering both system and input source variations, to generate the projection subspace. Such a reduction scheme is similar to the EKS/IEKS method [24, 9]. But the new method is based on more globally accurate balanced truncation reduction method instead of the locally accurate Krylov subspace method as in EKS/IEKS. As a result, it can reduce systems with many terminals like power grid networks while preserving variational information. After the reduction, Monte-Carlo based statistical simulation is performed on the reduced system and the statistical responses of the original systems are obtained thereafter. Similar to varPMTBR, the varETBR only requires the simulation of the reduced circuit using any existing transient analysis method. It is insensitive to the number of variables and variation ranges in terms of computing costs and accuracy, which makes it very general and scalable. Experimental results, on a number of the IBM benchmark circuits [15] up to 1.6 million nodes, show that the varETBR can be up to 4500*X* faster than the Monte-Carlo method and it is much more scalable than the StoEKS method [12]. varETBR can solve very large power grid networks with large number of random variables, large variation ranges and different variational distributions.

The rest of this paper is as follows: Section 2 presents the variational power grid models used in this paper. Section 3 reviews the extended Krylov subspace methods and fast balanced truncation methods. Our new variational analysis method varETBR is presented in Section 4. Section 5 shows the experimental results and

*This work is funded in part by NSF CAREER Award No. CCF-0448534, in part by NSF grant under No. OISE-0623038 and in part by the visiting scholar fund of ASIC & System State-Key-Lab of Fudan University.

978-1-4244-2749-9/09/$25.00 ©2009 IEEE

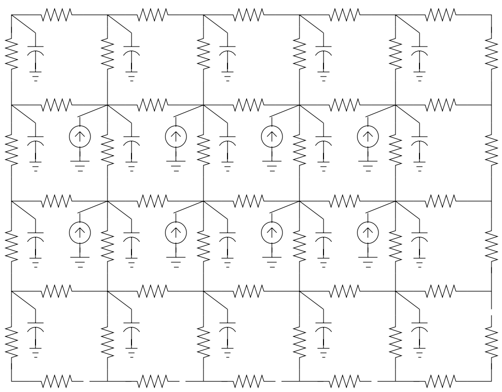


Figure 1: The power grid model used.

Section 6 concludes this paper.

2. POWER GRID NETWORK MODELS

2.1 Nominal Model

The power grid networks in this paper are modeled as RC networks with known time-variant current sources, which can be obtained by gate level logic simulations of the circuits. Fig. 1 shows the power grid models used in this paper. For a power grid (versus the ground grid), some nodes having known voltage are modeled as constant voltage sources. For C4 power grids, the known voltage nodes can be internal nodes inside the power grid. Given the current source vector, $u(t)$, the node voltages can be obtained by solving the following differential equations, which is formulated using modified nodal analysis (MNA) approach,

$$Gv(t) + C\frac{dv(t)}{dt} = Bu(t) \tag{1}$$

where $G \in R^{n\times n}$ is the conductance matrix, $C \in R^{n\times n}$ is the matrix resulting from storage elements. $v(t)$ is the vector of time-variant node voltages and branch currents of voltage sources. $u(t)$ is the vector of independent sources, and B is the input selector matrix.

2.2 Variational Model

In the presence of process variations, the G and C matrices and input currents $u(t)$ depend on variational circuit parameters, such as metal wire width, length, metal thickness on power grids, and transistor parameters, such as channel length, width, gate oxide thickness, etc. Process-induced random variations can be systemic and random and can be highly partially correlated [3]. For highly correlated variations like the inter-die variations, the worst case corner can be easily found by setting the parameters to their range limits (mean plus 3σ). The difficulty lies in the intra-die variations, where the circuit parameters are not correlated or spatially correlated. Intra-die variations also consist of local and layout dependent deterministic components and random components. In this paper, we focus on the random variations, which typically are modeled as multivariate Gaussian process with any spatial correlation [2].

We assume that we have a number of independent (uncorrelated) transformed orthonormal Gaussian random variables $\xi = [\xi_1, ..., \xi_M]$, which actually model the channel length, the device threshold voltage and the wire geometry variations. Therefore, the MNA equation for (1) becomes

$$G(\xi)v(t) + C(\xi)\frac{dv(t)}{dt} = Bu(t,\xi) \tag{2}$$

The spatial correlation in the intra-die variation can be processed by using the principal component analysis method (or other methods like K-L transformation or principal factor analysis, etc.) to transform the correlated variables into un-correlated variables before spectral statistical analysis [8].

Note that the input vector $u(t,\xi) = i(t,\xi) + u_0(t)$, where the current vector $i(t,\xi)$ follows the log-normal distribution and has both deterministic and random components, and the input voltage vector $u_0(t)$ is not effected by ξ. In this paper, we assume the dynamic currents (power) due to circuit switching are still modeled as deterministic currents, so we only consider the leakage variations as they are more significant owning to their log-normal distributions. Specifically, we expand the variational G and C around their mean values and keep the first order terms as in [11, 4, 18].

$$\begin{aligned} G(\xi) &= G_0 + G_1\xi_1 + G_2\xi_2 + ... + G_M\xi_M \\ C(\xi) &= C_0 + C_1\xi_1 + C_2\xi_2 + ... + C_M\xi_M \end{aligned} \tag{3}$$

We remark that the proposed method can be trivially extended to the second and higher order terms. The input current variation $i(t,\xi)$ follows the log-normal distribution as leakage variations are dominant factors:

$$i(\xi) = e^{g(\xi)}, \quad g(\xi) = \mu + \sigma\xi \tag{4}$$

Note that input current variation $i(\xi)$ is not function of time as we only model the static leakage variations for the simplicity of presentation. Our approach can be also applied to time-variant variations with any distribution.

3. EKS AND BALANCED FAST TRUNCATION METHODS

3.1 Extended Krylov subspace methods

Krylov subspace based model order reduction methods have been well-accepted in interconnect modeling. But the methods are less efficient for on-chip power supply network analysis, due to the presence of a large number of inputs (supply current excitations) and outputs (potential supply voltage degradation nodes). The main reason is that the computation costs and the projection vectors directly depend on the terminal count and the reduced model may increase very quickly with increasing terminal count.

To mitigate this problem, extended Krylov subspace method was proposed [24, 9]. The idea is to perform the reduction on both model and input in the moment form. As a result, the original multi-input and multi-output reduction problem becomes single-input and multi-output problem. One-side Krylov subspace method like PRIMA [16] can be efficiently used for reducing such systems. The computation now is independent of the number of terminals in the network. IEKS [9] shows that for piece-wise linear (PWL) sources, which is approximated by sums of delayed ramps in Laplace domain, the $1/s$ and $1/s^2$ terms are always zero. So no moment shifting is required as in [24].

Specifically, instead of computing the vectors of the n-th order moments for explicit moment matching, extended Krylov subspace method constructs a modified Krylov subspace by orthonormalizing the moment vectors with current sources. For a RLC network in frequency domain,

$$(G + sC)v(s) = Bu(s) \tag{5}$$

where G is a conductance matrix, C is a storage element matrix. B is a position matrix. If we expand $v(s)$ and $u(s)$ in moment form, we have

$$(G + sC)(m_0 + m_1 s + m_2 s^2 + ...) = B(u_0 + u_1 s + u_2 s^2 + ...) \tag{6}$$

The EKS algorithm essentially performs the orthonormalization on the response moments m_i. After we obtain the reduced system, we can perform the transient simulation on the reduced system,

$$\hat{G}\hat{v}(t)+\hat{C}\frac{d\hat{v}(t)}{dt}=\hat{B}u(t) \tag{7}$$

Transient simulation can be carried out on (7), which will be very efficient due to reduced circuit matrices. After this, the original waveforms can be obtained by $v(t) = V\hat{v}(t)$.

3.2 Fast TBR method: Poor man's TBR

In truncated balanced realization (TBR) based reduction method, there are two steps in the reduction process: The *balancing* step transforms the states such that states can be controlled and observed equally. The *truncating* step then throws away the weak states, which usually leads to much smaller models. The major advantage of TBR methods is that TBR methods can give a deterministic global bound for the approximate error and it can give nearly optimal models in terms of errors and model sizes.

The TBR method in general suffers high computation costs, as it needs to solve expensive the Lyapunov equation (8) as shown below. To mitigate this problem, the fast TBR method, called Poor man's TBR or PMTBR [19] was proposed, which computes the approximate Grammians. Specifically, for a symmetric circuit matrices with $A = A^T$ and $C = B^T$, which is the case for RC and RL circuits. It is easy to see that, both Grammians are equal and can be obtained by solving the Lyapunov equation:

$$AX+XA^T+BB^T=0 \tag{8}$$

Since X is symmetric, it is orthogonally diagonalizable, i.e., there exists $T^{-1} = T^T$ such that $T^TXT = \Sigma$. Then, we have

$$T^TXXT=(T^TXT)(T^TXT)=(\Sigma)^2 \tag{9}$$

which means, in this symmetrized case, the eigenspace of Grammian product XX is exactly the eigenspace of each Grammian X and we only need to find the dominant invariant subspace of an approximated Grammian $\hat{X}$. In frequency domain, the Grammian X can also be computed from the expression

$$X=\int_{-\infty}^{+\infty}(j\omega I-A)^{-1}BB^T(j\omega I-A)^{-H}d\omega \tag{10}$$

where superscript H denotes Hermitian transpose. Let ω_k be kth sampling point. If we define

$$z_k=(j\omega_k I-A)^{-1}B \tag{11}$$

then based on the numerical quadrature rule, X can be approximated as

$$\hat{X}=\sum w_k z_k z_k^H=ZW^2Z^H \tag{12}$$

where $Z = [z_1, z_2, \ldots, z_n]$. W a diagonal matrix with diagonal entries $w_{kk} = \sqrt{w_k}$. w_k comes from a specific numerical quadrature method. Since $\hat{X}$ is symmetric, it is orthogonally diagonalizable.

$$\hat{V}^T\hat{X}\hat{V}=\begin{bmatrix}\hat{V}_1^T\\ \hat{V}_2^T\end{bmatrix}\hat{X}\begin{bmatrix}\hat{V}_1 & \hat{V}_2\end{bmatrix}=\begin{bmatrix}\hat{\Sigma}_1 & 0\\ 0 & \hat{\Sigma}_2\end{bmatrix} \tag{13}$$

where $\hat{V}^T\hat{V} = I$. $\hat{V}$ converges to the eigenspaces of X and the dominant eigenvectors $\hat{V}_1$ can be used as the projection matrix in a model reduction approach ($A_r = \hat{V}_1^T A\hat{V}_1, B_r = \hat{V}_1^T B$).

3.3 Variational Poor man's TBR

In [18], PMTBR has been extended to reduce interconnect circuits with parameter variations. Notice that PMTBR can be viewed as one-dimensional numerical quadrature with respect to complex frequency s . The varPMTBR method is multi-dimensional numerical quadrature with respect to random variables, in addition to the complex frequency variables s in PMTBR. As a result, a variational Grammian is defined, which is the mean of Grammians over all the variational variables and is approximated by the Monte-Carlo method.

One important observation in varPMTBR is that number of samplings in building subspaces are much smaller than the number of general Monte-Carlo samplings at the circuit level to achieve the same accuracy. As a result, varPMTMR is much faster than the brute-force Monte-Carlo method and its costs are much less sensitive to the number of random variables and variation ranges, which makes this method much more efficient than the existing variational or parameterized model order reduction methods [25]. The algorithm flow is shown in *Algorithm 1* below:

Algorithm 1: varPMTBR: Variational Truncated Balanced Realization method

Input: Circuit of $G(\xi)$, $C(\xi)$, B, variables $\xi = [\xi_1, ..., \xi_M]$, number of samples: q
Output: Reduced system matrices $\hat{G}(\xi)$, $\hat{C}(\xi)$, $\hat{B}$

1. Select q points over an M+1 dimensional space $(s, \xi_1, ..., \xi_M)$
2. Compute $z_k^r = (s_k C(\xi_1^k, ..., \xi_M^k) + G(\xi_1^k, ..., \xi_M^k))^{-1}B$
3. Form the matrix $Z_r = [z_1^r, z_2^r, \ldots, z_q^r]$
4. Perform SVD on Z_r, $Z_r = V_r S_r U_r^T$
5. $\hat{G}(\xi) = V_r^T G(\xi)V_r$,$\hat{C}(\xi) = V_r^T C(\xi)V_r$, $\hat{B} = V_r^T B$
6. End

4. NEW VARIATIONAL ANALYSIS METHOD: VARETBR

The new variational power grid analysis algorithm, varETBR, follows the similar spirit of EKS method [24, 9]. But it different from the EKS method substantially in the following aspects. First, it applies fast, more globally accurate, balanced truncation method to reduce the system considering input sources. Second, it employs n-dimentional numerical quadrature approximation to perform variational reduction.

4.1 Extended truncated balanced realization scheme

The new method is based on the recently proposed extended truncated balanced realization method [10]. We first review this method.

For a linear system in (1), we first define the response Grammian in the frequency domain as:

$$X_r=\int_{-\infty}^{+\infty}(j\omega C+G)^{-1}Bu(j\omega)u^T(j\omega)B^T(j\omega C+G)^{-H}d\omega \tag{14}$$

To fast compute the response Grammian X_r, we can follow the similar strategy in PMTBR method. Specifically, let ω_k be kth sampling point over the frequency range. If we further define

$$z_k^r=(j\omega_k C+G)^{-1}Bu(j\omega_k) \tag{15}$$

then $\hat{X}$ can be computed approximately by numerical quadrature methods

$$\hat{X}_r=\sum_k w_k z_k^r z_k^{rH}=Z_rW^2Z_r^H \tag{16}$$

where Z_r is a matrix whose columns are z_k^r and W a diagonal matrix with diagonal entries $w_{kk} = \sqrt{w_k}$. w_k comes from a specific quadrature method.

The projection matrix can be obtained by singular value decomposition of Z_r. After this, we can reduce the original matrices into small ones and then perform the transient analysis on the reduced circuit matrices. Also we find that weights w_k are not important for the SVD process. The weight matrix W will not change the subspace of Z_r as it simple multiplies each vector in Z_r with a constant. In our algorithm, we just simple ignore the weights. By using balanced truncation reduction, we can easily extend the resulting algorithm to perform variational analysis as shown in next subsection.

Notice that we need the frequency response caused by input signal $u(j\omega_k)$ in (15). This can be obtained by fast Fourier transformation on the input signals in time domain. Using frequency spectrum representations for the input signals is a significant improvement over the EKS method as we avoid the explicit moment representation of the current sources, which are not accurate for currents rich in high frequency components due to the well-known problems in explicit moment matching methods [21].

Another improvements is better accuracy owning to the use of the fast balanced truncation method for the reduction, which has global accuracy [13, 20]. The extended TBR algorithm is summarized in *Algorithm* 2. After the reduction, the reduced system in (7) can be

Algorithm 2: ETBR: Extended Truncated Balanced Realization method

---

Input: Circuit of G, C, B, $u(t)$, number of samples: q
Output: Reduced system matrices $\hat{G}$, $\hat{C}$, $\hat{B}$

---

1. Convert all the input signals $u(t)$ into $u(s)$ using FFT.
2. Select q frequency points $s_1, s_2, \ldots, s_q$ over the frequency range
3. Compute $z_k^r = (s_k C + G)^{-1} B u(s_k)$
4. Form the matrix $Z_r = [z_1^r, z_2^r, \ldots, z_q^r]$
5. Perform SVD on Z_r, $Z_r = V_r S_r U_r^T$
6. $\hat{G} = V_r^T G V_r$, $\hat{C} = V_r^T C V_r$, $\hat{B} = V_r^T B$
7. End

---

simulated in time domain and the original waveforms can be obtained by $v(t) = V_r \hat{v}(t)$. Note that, like the EKS method, we use congruence transformation for the reduction process with orthogonal columns in the projection matrix (by using Arnoldi or Arnoldi-like process), the reduced system must be stable. As far as simulation is concerned, this is good enough. If all the observable ports are also the current source nodes, i.e. $y(t) = B^T v(t)$, where $y(t)$ is the voltage vector at all observable ports, the reduced system is passive.

It was shown in [20] that the fast TBR method has the similar time complexity of the multiple-point Krylov subspace based reduction methods. The extended TBR method also has the similar computation costs of the EKS method.

4.2 The new varETBR method

To consider the process variations, we end up with new *variational Grammian*, where in addition to the frequency, there are other variables. Specifically, we first define the variational response Grammian in the frequency domain as following:

$$X_{r\xi} = E\{\int_{-\infty}^{+\infty} (j\omega C_\xi + G_\xi)^{-1} B u_\xi(j\omega) u_\xi^T(j\omega) B^T (j\omega C_\xi + G_\xi)^{-H} d\omega\} \tag{17}$$

The variational response Grammian essentially requires to compute high-dimentional integration, which can be numerically computed by the Monte-Carlo method. The varETBR algorithm flow is show in *Algorithm* 3.

The algorithm starts with the given a power grid network and the number of samplings q, are used for building the projection subspace. Then it computes the variational response $z_k^r = (s_k C(\xi_1^k, \ldots, \xi_M^k) + G(\xi_1^k, \ldots, \xi_M^k))^{-1} B u(s_k, \xi_1^k, \ldots, \xi_M^k))$. Then we perform the SVD on $Z_r = [z_1^r, z_2^r, \ldots, z_q^r]$ to construct the projection matrix. After the reduction, we perform the Monte-Carlo based statistical analysis to obtain the variational responses from $v(t) = V_r \hat{v}(t)$.

Algorithm 3: varETBR: Variational extended Truncated Balanced Realization method

---

Input: Circuit of $G(\xi)$, $C(\xi)$, B, $u(t,\xi)$, variables $\xi = [\xi_1, \ldots, \xi_M]$, number of samples: q
Output: The variational response $v(t)$

---

1. Convert all the nominal input signals $u(t)$ into $u(s)$ using FFT.
2. Select q points over an M+1 dimensional space $(s, \xi_1, \ldots, \xi_M)$
3. Compute $z_k^r = (s_k C(\xi_1^k, \ldots, \xi_M^k) + G(\xi_1^k, \ldots, \xi_M^k))^{-1} B u(s_k, \xi_1^k, \ldots, \xi_M^k))$
4. Form the matrix $Z_r = [z_1^r, z_2^r, \ldots, z_q^r]$
5. Perform SVD on Z_r, $Z_r = V_r S_r U_r^T$
6. $\hat{G} = V_r^T G(\xi) V_r$, $\hat{C} = V_r^T C(\xi) V_r$, $\hat{B} = V_r^T B$
7. Perform the Monte-Carlo simulation on $\hat{G}(\xi)\hat{v}(t) + \hat{C}(\xi)\frac{d\hat{v}(t)}{dt} = \hat{B}u(t,\xi)$
8. Obtain the variational response $v(t) = V_r \hat{v}(t)$.
9. End

Compared with the existing approaches, varETBR has the following advantages and features:

1. Only Monte-Carlo sampling method is used, varETBR is easy to implement and is very general for dealing with different variation distributions and large variation ranges.

2. Vary scalable for solving large networks with large number of variables as reduction is performed.

3. More amanable for parallel computing as each sampling in freqency domain can be done in parallel.

4. More accurate over wide band frequency ranges due to the global error bound provided by the TBR based methods (compared with locally accurate EKS method).

5. Avoid the explicit moment representation of the input signals, which can lead more accurate results than the EKS method when signals are rich in high frequency components.

5. EXPERIMENTAL RESULTS

The proposed *varETBR* algorithm has been implemented using MATLAB and tested on an Intel quad-core workstation with 16GB memory under Linux environment.

All the benchmarks are real PG circuits from IBM provided by [15]. But the circuits in [15] are resistor-only circuits. For transient analysis, we need to add capacitors and transient input waveforms. As a result, we modified the benchmark circuits. First we added one grounded capacitor on each node with a random value in the magnitude of pF. Second we replaced the DC current sources by a piecewise linear signal in the benchmark. The values of these signals are also randomly generated based on their original values in the DC benchmarks. We implemented a parser using Python to transform the SPICE format benchmarks into MATLAB format.

The summary of our transient PG benchmarks are show in Table 1. We use MNA formulation to set up the circuit matrices. To efficiently solve PG circuits with 1.6 million nodes in MATLAB, an external linear solver package UMFPACK [1] is used, which is linked with Matlab using Matlab mexFunction.

In sequel, we will compare varETBR with the Monte-Carlo method, first in accuracy and then in CPU times. In all the test cases, the number of samples used for forming the subspace in varETBR are 50 based on our experience. The reduced order is set to $p = 10$, which is accurate enough in practice. Here we set variation range to 10% and number of variables to 6 (2 for G, 2 for C and 2 for i). $G(\xi)$ and $C(\xi)$ follow Gaussian distribution. $i(t,\xi)$, which model the leakage variations [5], follows log-normal distribution.

Our varETBR is essentially a kind of reduced Monte-Carlo methods. It inherits the merits of Monte-Carlo methods, which is less

Table 1: PG benchmarks

Name	#Nodes	#V Sources	#I Sources
ibmpg1	30638	14308	10774
ibmpg2	127238	330	37926
ibmpg3	851584	955	201054
ibmpg4	953583	962	276976
ibmpg5	1079310	539087	540800
ibmpg6	1670494	836239	761484

sensitive to number of variables and can reflect the real distribution very accurately for sufficient number of samples. But the main disadvantage of Monte-Carlo is that it is too slow to simulate on large scale circuits. varETBR first reduces the size of circuits to a small number while maintaining the enough accuracy. And then varETBR can do Monte Carlo simulation on the reduced circuits very fast. Note that the reduction process is done only once during the simulation process.

To verify the accuracy of our varETBR method, we show the results of simulations on *ibmpg1* (100 samples). Fig. 2 show the results of varETBR and the pure Monte-Carlo method at the 1000th node (named n1_20583_11663 in SPICE format) of *ibmpg1*. The circuit equations in Monte-Carlo are solved by MATLAB.

The absolute errors and relative errors of *ibmpg1* are shown in Fig. 3. We can briefly see that errors are very small and our varETBR is very accurate. Note that the errors are not only influenced by the variations but also depends on reduced order. To increase the accuracy, we may increase the reduced order. In our tests, we set the reduced order $p = 10$ for all the benchmarks.

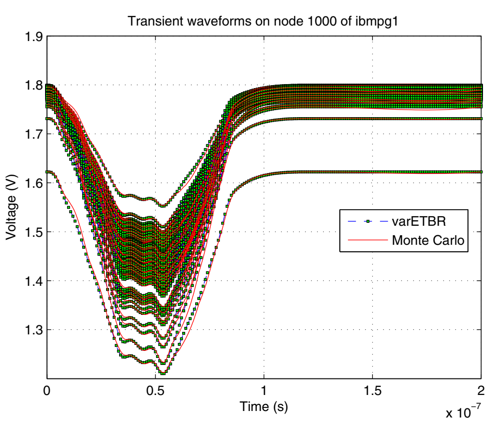
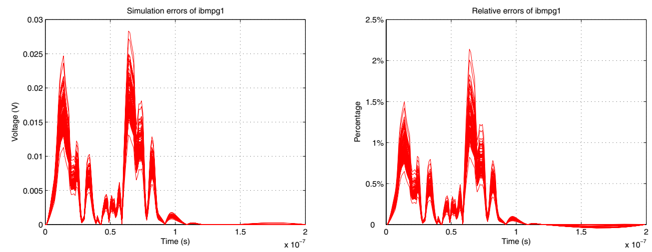


Figure 2: Transient waveform at the 1000th node (n1_20583_11663) of *ibmpg1* ($p = 10$, 100 samples).

Next we do accuracy comparison with Monte-Carlo for the probability distributions including means and variances. Fig. 4 shows the voltage distributions of both varETBR (blue, dash) and original Monte-Carlo (red, solid) at the 1000th node of *ibmpg1* when $t = 50ns$ (200 time steps between $0ns$ and $200ns$ in total). We can also refer to simulation waveforms on $t = 50ns$ in Fig. 2. Note that the results do not follow Gaussian distribution as $G(\xi)$ and $C(\xi)$ follow Gaussian distribution and $i(t, \xi)$ follows log-normal distribution. From Fig. 4, we can see that not only are the means and the variances of varETBR and Monte Carl almost the same, but also are their probability distributions.

Finally, we compare the CPU times of varETBR and the pure Monte-Carlo method. To verify the efficiency of our varETBR on both CPU time and memory, we do not need to run simulations



(a) Simulation errors of *ibmpg1* (100 samples). (b) Relative errors of *ibmpg1* (100 samples).

Figure 3: Simulation errors and relative errors of *ibmpg1*

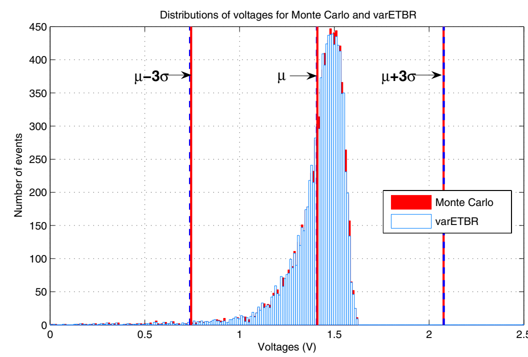


Figure 4: Voltage distribution at the 1000th node of *ibmpg1* (10000 samples) when $t = 50ns$.

Table 2: CPU times (s) comparison of varETBR and Monte-Carlo ($p = 10$)

Test Ckts	varETBR (s) Red. (s)	varETBR (s) Sim. (s)	Monte-Carlo Sim. (s)	Speedup
ibmpg1 (100)	23	14	739	53X
ibmpg1 (10000)	23	1335	70719	53X
ibmpg2 (10)	115	1.4	536	383X
ibmpg3 (10)	1879	1.5	4973	3315X
ibmpg4 (10)	2130	1.3	5275	4508X
ibmpg5 (10)	1439	1.3	5130	3946X
ibmpg6 (10)	1957	1.5	6774	4516X

Table 3: Relative errors for the mean of max voltage drop of varETBR compared with Monte-Carlo on the 2000th node of *ibmpg1* ($p = 10$, 10000 samples) for different variation ranges and different numbers of variables

#Variables	Variation $var = 10\%$	$var = 30\%$	$var = 50\%$	$var = 100\%$
$M = 6$	0.16%	0.08%	0.17%	0.21%
$M = 9$	0.16%	0.25%	0.08%	0.23%
$M = 12$	0.25%	0.07%	0.07%	0.28%
$M = 15$	0.15%	0.06%	0.05%	0.06%

many times for both varETBR and Monte-Carlo. We will run 10 or

Table 4: CPU times (s) comparison of StoEKS and varETBR (reduced order $p = 10$) with 10000 samples for different numbers of variables.

Test Ckts	M = 5		M = 7		M = 9	
	StoEKS (s)	varETBR (s)	StoEKS (s)	varETBR (s)	StoEKS (s)	varETBR (s)
ibmpg1	165	1315	572	1338	3748	1326
ibmpg2	1458	1387	–	1351	–	1377

100 samples for each benchmark to show the efficiency of varETBR since we already showed its accuracy. Although we only run small number of samples, the speedup will be the same.

Table 2 shows the CPU times of both varETBR (including FFT costs) and Monte-Carlo on the given set of circuits. The reduction order is $p = 10$. In varETBR, circuit model becomes much smaller after reduction and we only need to do the reduction once. Therefore, the simulation time is much faster than Monte-Carlo (up to 4500X). Basically, the bigger the original circuit size is, the faster the simulation will run for varETBR. For standard random simulation using random samples, the reduction time is negligible compared to the total simulation time. Note that we run random simulation 10000 times for *ibmpg1*, as shown in Table 2, to show the efficiency of our varETBR in practice.

It can be seen that varETBR is very scalable. It is in practice almost independent of the variation ranges and numbers of variables. One possible reason is that varETBR inherits the merits of random sampling in Monte-Carlo methods that is high scalability and robustness. When we increase the variation range and the number of variables, the accuracy of varETBR is almost unchanged. Table 3 shows that varETBR is very insensitive to the number of variables and variation ranges for a given circuit *ibmpg1*, where simulations are run on 10000 samples for both varETBR (reduced order $p = 10$) and Monte-Carlo.

To demonstrate the efficiency of varETBR, we compare it with one recently proposed similar approach, *StoEKS* method, which employs Krylov subspace reduction with orthogonal polynomials in [12] on the same suite of IBM circuit. Table 4 shows the comparison results where '-' means out of memory error. StoEKS can only finish smaller circuits *ibmpg1* (30k) and *ibmpg2* (120k), while varETBR can go through all the benchmarks (up to 1.6M nodes) easily. The CPU time of StoEKS increases rapidly and even could not complete computations as variables count increases. For varETBR, CPU time is independent of number of variables and only depends on the reduced order and number of samples used in reduced Monte-Carlo simulation. Here we select reduced order $p = 10$ and 10000 samples that are enough in practice to obtain the accurate probability distribution.

6. CONCLUSION

In this paper, we have proposed a new scalable statistical power grid analysis approach based on extended truncated balanced realization reduction techniques. The new method, called *varETBR*, performs reduction on the original system using variation-bearing subspaces before Monte-Carlo statistical transient simulation. But different from the variational Poor man's TBR method, both system and input source variations are considered for generating the projection subspace by sampling variational response Gramians to perform the reduction. As a result, *varETBR* can reduce systems with many terminals like power grid networks while preserving variational information. After the reduction, Monte-Carlo based statistical simulation is performed on the reduced system to obtain the statistical responses of the original system. Experimental results show that the varETBR can be 4500X faster than the Monte-Carlo method and be scalable to solve very large power grid networks with large number of random variables and variation ranges. varETBR is also much more scalable than the StoEKS [12] on the IBM benchmark circuits.

7. REFERENCES

[1] Umfpack. http://www.cise.ufl.edu/research/sparse/umfpack/.

[2] A. B.Kahng. DFM tools and methodologies for 65nm and below. In *Proc. Asia South Pacific Design Automation Conf. (ASPDAC)*, 2006. Tutorial.

[3] C. Chiang and J. Kawa. *Design for Manufacturability*. Springer, 2007.

[4] L. Daniel, O. C. Siong, L. S. Chay, K. H. Lee, and J. White. Multi-parameter moment-matching model-reduction approach for generating geometrically parameterized interconnect performance models. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 23(5):678–693, May 2004.

[5] I. A. Ferzli and F. N. Najm. Statistical estimation of leakage-induced power grid voltage drop considering within-die process variations. In *Proc. Design Automation Conf. (DAC)*, pages 865–859, 2003.

[6] I. A. Ferzli and F. N. Najm. Statistical verification of power grids considering process-induced leakage current variations. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 770–777, 2003.

[7] P. Ghanta, S. Vrudhula, and S. Bhardwaj. Stochasic variational analysis of large power grids considering intra-die correlations. In *Proc. Design Automation Conf. (DAC)*, pages 211–216, July 2006.

[8] P. Ghanta, S. Vrudhula, R. Panda, and J. Wang. Stochastic power grid analysis considering process variations. In *Proc. European Design and Test Conf. (DATE)*, volume 2, pages 964–969, 2005.

[9] Y. Lee, Y. Cao, T. Chen, J. Wang, and C. Chen. HiPRIME: Hierarchical and passivity preserved interconnect macromodeling engine for rlkc power delivery. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 24(6):797–806, 2005.

[10] D. Li, S. X.-D. Tan, and B. McGaughy. ETBR: Extended truncated balanced realization method for on-chip power grid network analysis. In *Proc. European Design and Test Conf. (DATE)*, pages 432–437, 2008.

[11] Y. Liu, L. T. Pileggi, and A. J. Strojwas. Model order-reduction of rc(l) interconnect including variational analysis. In *DAC '99: Proceedings of the 36th ACM/IEEE conference on Design automation*, pages 201–206, 1999.

[12] N. Mi, S. X.-D. Tan, P. Liu, J. Cui, Y. Cai, and X. Hong. Stochastic extended Krylov subspace method for variational analysis of on-chip power grid networks. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 48–53, 2007.

[13] B. Moore. Principle component analysis in linear systems: Controllability, and observability, and model reduction. *IEEE Trans. Automat. Contr.*, 26(1):17–32, 1981.

[14] S. Nassif. Delay variability: sources, impact and trends. In *Proc. IEEE Int. Solid-State Circuits Conf.*, pages 368–369, San Francisco, CA, Feb 2000.

[15] S. R. Nassif. Power grid analysis benchmarks. In *Proc. Asia South Pacific Design Automation Conf. (ASPDAC)*, pages 376–381, 2008.

[16] A. Odabasioglu, M. Celik, and L. Pileggi. PRIMA: Passive reduced-order interconnect macromodeling algorithm. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, pages 645–654, 1998.

[17] S. Pant, D. Blaauw, V. Zolotov, S. Sundareswaran, and R. Panda. A stochastic approach to power grid analysis. In *Proc. Design Automation Conf. (DAC)*, pages 171–176, 2004.

[18] J. Phillips. Variational interconnect analysis via PMTBR. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 872–879, Nov. 2004.

[19] J. R. Phillips and L. M. Silveira. Poor man's TBR: a simple model reduction scheme. In *Proc. European Design and Test Conf. (DATE)*, pages 938–943, 2004.

[20] J. R. Phillips and L. M. Silveira. Poor man's TBR: a simple model reduction scheme. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 24(1):43– 55, 2005.

[21] L. T. Pillage, R. A. Rohrer, and C. Visweswariah. *Electronic Circuit and System Simulation Methods*. McGraw-Hill, New York, 1994.

[22] R. Rutenbar. Next-generation design and EDA challenges. In *Proc. Asia South Pacific Design Automation Conf. (ASPDAC)*, January 2007. Keynote speech.

[23] A. Srivastava, R. Bai, D. Blaauw, and D. Sylvester. Modeling and analysis of leakage power considering within-die process variations. In *Proc. Int. Symp. on Low Power Electronics and Design(ISLPED)*, pages 64–67, Aug. 2002.

[24] J. M. Wang and T. V. Nguyen. Extended Krylov subspace method for reduced order analysis of linear circuit with multiple sources. In *Proc. Design Automation Conf. (DAC)*, pages 247–252, 2000.

[25] Z. Zhu and J. Phillips. Random sampling of moment graph: a stochastic Krylov-reduction algorithm. In *Proc. European Design and Test Conf. (DATE)*, pages 1502–1507, April 2007.