Hierarchical Simulation of Process Variations and Their Impact on Circuits and Systems: Methodology

Jürgen K. Lorenz, *Member, IEEE*, Eberhard Bär, Tanja Clees, Roland Jancke, Christian P. J. Salzig, and Siegfried Selberherr, *Fellow, IEEE*

***Abstract*—Process variations increasingly challenge the manufacturability of advanced devices and the yield of integrated circuits. Technology computer-aided design (TCAD) has the potential to make key contributions to minimize this problem, by assessing the impact of certain variations on the device, circuit, and system. In this way, TCAD can provide the information necessary to decide on investments in the processing level or the adoption of a more variation tolerant process flow, device architecture, or design on circuit or chip level. In this first of two consecutive papers, sources of process variations and the state of the art of related simulation tools are reviewed. An approach for hierarchical simulation of process variations including their correlations is presented. The second paper, also published in this issue, presents examples of simulation results obtained with this methodology.**

***Index Terms*—Circuit simulation, manufacturability, process modeling, semiconductor device modeling, sensitivity, yield.**

I. Introduction

PROCESS variations increasingly affect the performance, the reliability, and the manufacturability of advanced semiconductor devices. This is among others highlighted in the International Technology Roadmap for Semiconductors (ITRS) [1], where the problem of variability is addressed in many and quite different chapters, ranging from design through lithography and front-end processes to metrology. In the *Design* chapter, "variability and lithography limitations" are a key concern of the "Design for Manufacturability" section. Specifications for 3σ variations are particularly given for threshold voltage and critical dimensions (CDs). Aside from the patterning of minimum feature sizes by lithography and etching, the 3σ variations of the created CDs are essential requirements in the *Lithography* and *Front End Processes* chapters. For etching, gate-length 3σ values are even separately specified across the chip, across the wafer, between wafers, and lot to lot. Even in the *Metrology* chapter, variability is mentioned as one of the key problems because of measurement uncertainty results from three sources, i.e., reproducibility of measurement of the same sample with the same tool, differences of measuring the same sample with different tools, *and variations between samples*. These examples of the broad discussion of variability throughout the ITRS stress its importance for current and future advanced scaled devices.

Simulation is the only realistic possibility for a holistic investigation of the impact of the large manifold of relevant process variations on circuits and systems. However, while, in process and device simulations, up to some million variables are used to describe one device due to the necessity to express distributions of physical quantities in space and time, circuits and systems are usually described with a relatively small number of parameters in compact or behavioral models. Therefore, a hierarchical approach is required to investigate process variations from their sources, which are largely at the equipment level, to their impact on devices, circuits, and systems. Simulation tools are needed, which accurately predict not only nominal values but also their variability, both along the chain of technological processes and at all levels. A key requirement is the appropriate transfer of data between different tools and/or levels, which makes sure that variability is not mixed up with, e.g., discretization errors of the algorithms applied. Advancing from one level to the other, particularly from the nanoscale simulation of processes and devices with spatially resolved meshes to compact device models, requires suitable data reduction, which maintains all information needed at later steps of the simulation sequence, including variations.

In the following, the most relevant sources of process variations and the state of the art of simulation tools are discussed. An approach for the hierarchical simulation of variations including their correlations, developed at Fraunhofer, is presented. Application results are shown in another paper included in this Special Issue [2].

II. Sources of Process Variations

There are basically two different kinds of process variations, i.e., those due to material properties including particle statistics and those due to equipment issues.

Most frequently discussed in the literature are statistical variations that are due to the atomistic or molecular nature of

Manuscript received November 16, 2010; revised February 23, 2011; accepted March 31, 2011. Date of publication June 16, 2011; date of current version July 22, 2011. This work was supported in part by the Fraunhofer Internal Programs under Grant MAVO 817 759. The review of this paper was arranged by Editor A. Asenov.

J. K. Lorenz and E. Bär are with the Fraunhofer Institute for Integrated Systems and Device Technology (IISB), 91058 Erlangen, Germany (e-mail: juergen.lorenz@iisb.fraunhofer.de; eberhard.baer@iisb.fraunhofer.de).

T. Clees is with the Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), 53754 Sankt Augustin, Germany (e-mail: tanja.clees@scai.fraunhofer.de).

R. Jancke is with the Division Design Automation of the Fraunhofer Institute for Integrated Circuits (IIS/EAS), 01069 Dresden, Germany (e-mail: roland.jancke@eas.iis.fraunhofer.de).

C. P. J. Salzig is with the Fraunhofer Institute for Industrial Mathematics (ITWM), 67663 Kaiserslautern, Germany (e-mail: christian.salzig@itwm.fraunhofer.de).

S. Selberherr is with the Institute for Microelectronics of the Technical University Vienna, 1040 Vienna, Austria (e-mail: Selberherr@TUWien.ac.at).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TED.2011.2150225

0018-9383/$26.00 © 2011 IEEE

some processes. In particular, low-dose implantations into very small areas result in an average overall amount of just a few implanted atoms and, in turn, considerable fluctuations around the desired nominal value. Since this is a purely statistical effect, a particular process simulation is not required to describe it, but the Monte Carlo simulation of ion implantation can be used to visualize it. The impact of this effect on devices and simple circuits has been discussed in numerous papers, e.g., [3] and [4], using properly extended device simulation programs and simple but justified assumptions on the ion statistics. Quite similar approaches have been used for the simulation of the impact of line-edge and line-width roughness (LER/LWR) [5], which is a statistical deviation from the ideal straight-line-shape-patterned structures. Moreover, defects (particles) statistically distributed on a mask may print on the photoresist in lithography steps and therefore lead to variations. For silicon-on-insulator (SOI) devices, variations of the thicknesses of the silicon layer contribute to the statistical variations. For advanced SOI substrates, this variation across the wafer is less than 1 nm, and empirically, an impact of about 25 mV/nm was found [6]. Own studies for fully depleted (FD) SOI transistors based on ITRS specifications for silicon film thickness variations led to similar figures [7]. An important common feature of all these statistical variations is that their appearance on different dies, across the wafer, and on different wafers is statistically independent. In turn, they cannot be reduced or compensated by advanced process control methods, which are today commonly used in manufacturing to increase yield.

Lesser attention has been paid in research to systematic variations that result from the fabrication equipment and its use or are due to effects of the wafer itself or its patterning. These variations are exactly the main subject of this paper and the consecutive paper [2].

Key sources of systematic variations are the lithography steps used to pattern the wafers. Currently, optical lithography is being driven to its physical limits, particularly because the lack of lens materials suitable for wavelengths below 193 nm rules out a further reduction of the wavelength in manufacturing. Therefore, the use of tricks is requested, and sophisticated methods are applied, such as off-axis and/or incoherent illumination, phase-shifting masks, immersion lithography, and, recently, double/multiple patterning. LWR/LER has been previously mentioned as a statistical variation that is generated during lithography steps due to the molecular nature of the photoresist. However, its size also depends on the aerial image quality and several resist quantities, and can be simulated with mesoscopic models [8]. Aside from this, several systematic variations occur in lithography: Any stepper causes variations of the focus position of the optical system versus the photoresist, due to variations of the vertical distance of the stepper from the wafer, wafer warpage, topography of wafers that already underwent processing steps, and changes in resist thickness. Misalignment between different lithography steps is a premier source of variations. Aside from these systematic variations, the statistical spread of the laser power, which leads to fluctuations of the illumination dose, also contributes to variability in lithography. Further sources of variations are lens aberrations and mask imperfections—their imaging on the wafer is expressed by the mask error enhancement factor (MEEF), which is a critical quantity in advanced lithography. These variations partly occur between different wafers only, partly between different exposure fields on the same wafer, partly (wafer topography and warpage) even within the die. The situation gets worse in the case of double patterning steps, where two mask levels are used to pattern minimum-size features, which makes even adjacent patterns subject to all these variations. Another key problem is that even homogeneous or symmetric variations of lithography parameters, e.g., defocus, result in highly inhomogeneous or asymmetric variations of the patterned features. The reason is shown in the process window that is well known in lithography (see Fig. 1). Here, the printed CD is sketched versus the defocus and the relative intensity. The middle line represents the nominal CD, whereas the lower and upper lines represent an increase and a decrease in the CD by 10%, respectively. In consequence, deviations from the focus position in either direction result in the same sign of CD variations, giving rise to highly asymmetric distributions, as presented in [2, Sec. III]. Not discussed here is the optical proximity effect: Via diffraction in imaging, structures influence each other up to a distance of some wavelengths. This effect is to some extent already being taken into account in the optical proximity correction (OPC) used in state-of-the-art design tools. The use of extreme-ultraviolet (EUV) lithography would change many details but not the basic problem.

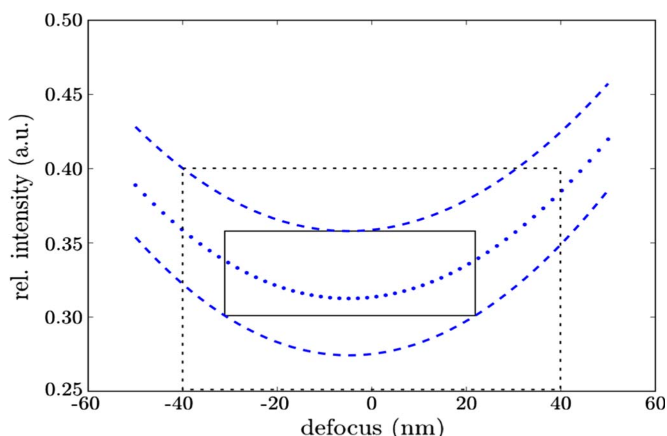

Fig. 1. Typical aerial-image-based process window in optical lithography: Defocus and relative intensity range for lines with nominal and ±10% CDs.

Several sources of across-wafer and wafer-to-wafer variations must be considered in other topography processes such as the following: in deposition and etching, spatial distribution of neutrals and ions within the plasma, inhomogeneities of the gas flow, and variations of temperature and pressure; in sputter deposition, spatial variations of target erosion and target aging; and in chemical mechanical polishing (CMP), pad pressure variations and slurry flow distribution, pad aging, and variations of the chemical composition of the slurry. The treatment of these variations requires appropriate equipment simulation and subsequent data transfer to feature-scale (process) simulation. Moreover, these processes are also affected by pattern-dependent effects that lead to the following intra-die variations: aspect-ratio dependence and iso-dense bias in etching; influence of aspect-ratio-dependent shadowing on deposition

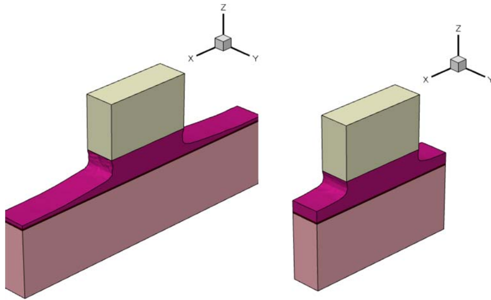

Fig. 2. Example for the effect of aspect-ratio-dependent etching: (Left) Isolated line versus (right) dense line. The layers from top to bottom are resist, polysilicon, oxide, and the silicon substrate.

rates inside features; and dependence of polishing rates and feature profiles (dishing and erosion) on pattern density in CMP processes. Fig. 2 shows an example for aspect-ratio-dependent etching.

Ion implantation, in the past, could be considered as a well-controlled process because of the separation of implanted ions and their energies, a high number of ions being implanted into the relevant areas, and the minimization of channeling by suitable tilt and rotation angles, thereby also minimizing the impact of angle variations during scanning. However, current low-energy processes lead to broadening of the energy distribution of the ions due to deceleration. The impact of variations of tilt and rotation angles becomes more severe due to the larger residual channeling at lower energies. This effect becomes even more pronounced for plasma immersion doping, due to its inherent larger angular spread. Dopant fluctuations resulting from ion implantation were already previously discussed.

In traditional furnace annealing and oxidation processes, temperature variations due to inhomogeneities in heating and gas flow lead to across-wafer and inter-wafer variations. Furthermore, temporal and spatial variations in temperature profiles can have a similar effect. Current nonmelting laser and Flash annealing processes are much more critical in terms of variations because process times and, therefore, thermal diffusion lengths are too short to smear out inhomogeneous heat transfer into the wafer caused by patterned structures with changes in wafer reflectivity. In consequence, not only temperature changes with time must be considered but also spatial variations. This can lead to intra-die variations of dopant activation.

In addition to the pattern effects previously mentioned for most processing steps, pattern-dependent stress is also a major source of intra-die variability. It not only directly influences processing steps such as oxidation but also particularly affects carrier mobility. Other main sources of variability not previously mentioned so far are interface charges and traps, and surface roughness in the channel, which affect carrier mobility and, partly, threshold voltage. Interconnect surface roughness affects, in particular, resistivity and electromigration.

Compared with variability sources such as dopant fluctuations, LER/LWR, and interface traps, which have been frequently discussed in the literature, e.g., [3]–[5], the process variations previously summarized, which affect device geometry and dopant activation, have received lesser attention so far. One reason seems to be that research projects on the development of advanced device architectures mostly apply e-beam direct WRITE for pattern generation. Moreover, for bulk metal–oxide–semiconductor field-effect transistor (MOSFET) devices, the large effect of dopant fluctuations may hide other variations. However, the situation completely changes in the case of undoped channels, e.g., in FD SOI transistors. Frequently, variations at the process level depend not only on the equipment but also on patterns already existent on the wafer and must be therefore quantified on a case-by-case basis.

III. State of the-Art in Simulation of Variability

This section focuses on the most important aspects for the treatment of process variations at the various levels of simulation, whereas discussions of the general models and algorithms used for the simulation of nominal processes are beyond the scope of this paper and were published elsewhere (see, e.g., [9] for the modeling of diffusion and activation).

A. Equipment Simulation

Within micro- and nanoelectronics, equipment simulation is mostly used to study reactors for etching and deposition. For this purpose, fluid dynamics or multiphysics codes are used, which were extended to include energy transfer into the equipment, e.g., by radiative or inductive heating, plasma dynamics, and particularly the chemical reactions that govern the deposition or etching process. A typical example is the program CFD-ACE [10], which has been also applied for the etching simulations on the equipment scale reported below. From the simulation of gas flow, energy transfer, plasma dynamics, and chemical reactions, the distributions of temperatures and various neutral and ionic species throughout the reactor chamber are simulated. The geometry to be considered includes not only the process chamber of the reactor with its gas inlets and outlets and its heating or inductive elements but also an approximation of the semiconductor wafers to be processed. The distributions of temperature, concentrations, energies, and directions of the reactive particles simulated at the position of the wafer can be subsequently used to extract etching and/or deposition rates at the wafer surface and changes across a wafer or between different wafers. They can be also used in subsequent feature-scale simulations, as summarized below. Various implementations have been reported in the literature, e.g., [11].

Concerning diffusion and oxidation furnaces, activities have shifted from the simulation of temperature distributions and partly gas compositions in (horizontal) batch reactors to the investigation of contemporary millisecond or Flash annealing reactors. Here, the main problem is that, increasingly, the spatial and temporal distributions of the temperature at the wafer are difficult to characterize and are partly even not fully reproducible. However, for the simulation of the activation and the diffusion of dopant atoms, not only the nominal temperature is important but also its distribution, which additionally also depends on wafer reflectivity and, in turn, on layer stacks and pattern densities. While the study of such effects was already

started in the 1980s [12], much work still remains to be carried out on the characterization and the simulation of the current state-of-the-art equipment.

B. Lithography Simulation

The situation of lithography is insofar specific as no reasonable separation can be made between equipment and process simulation. The main aspect to be simulated is first the deposition of light intensity into the photoresist (illumination), which is determined by the diffraction of the illumination light at the optical mask and the transformation of the mask diffraction spectrum into the wafer stack by the projection system. Then, the diffusion and reaction processes in the photoresist (development) must be simulated, which finally lead to the dissolution of the resist in the sufficiently exposed or unexposed resist areas, for positive- or negative-tone resists, respectively. While models and parameters for resist development generally still lack predictive accuracy and, in turn, require a significant amount of calibration, the optical imaging process is accurately described by Maxwell's equations. Particularly in lithography, rigorous three-dimensional (3-D) simulation is mandatory for almost all relevant cases. Commercial tools available for the simulation of optical lithography, such as *PROLITH* [13], *Panoramic* [14], and *Sentaurus Lithography* [15], employ very different approaches to solve Maxwell's equations or to partly replace them by simpler and less time-consuming methods. Furthermore, they largely differ in the approximations used, e.g., for the frequently used off-axis illumination ("Hopkins" or "non-Hopkins" approach [16]) or for nonpoint light sources. *Optolith* [17] is even only applicable for 2-D problems. For this paper and the consecutive paper [2], the proprietary 3-D research lithography simulator *Dr.LiTHO* [18] from the Fraunhofer IISB has been extended, integrated and applied. This tool allows for the accurate and computationally efficient solution of Maxwell's equations by sophisticated waveguide and domain decomposition methods [19] and has been combined with a genetic algorithm e.g., for the cooptimization of light sources and masks in order to optimize not only the nominal features printed but also the process windows obtained, thus minimizing the effect of lithography parameter variations, as summarized in Section II.

C. Process Simulation

Sophisticated feature-scale models for the simulation of ion implantation, annealing, and oxidation are implemented in commercial tools such as *Sentaurus Process* [15] or *Victory* [17] and are subject of further research in various projects such as the current European Commission project Advanced TEchnology MOdeling for eXtra-functionality devices (ATEMOX) [20], in which the generality and the predictive accuracy of the models are being further improved. The most important limiting factors for the investigation of process variations are, particularly for 3-D simulations, the long computation times needed and the still-not-fully-solved problem of generating and adapting numerical meshes for nonplanar moving geometries in three dimensions, particularly for oxidation steps.

The capabilities for simulation of etching and deposition processes in these tools is mostly limited to 2-D and/or simple solid modeling approaches to create the input geometries for 3-D device simulation, e.g., by extruding 2-D geometries to 3-D. Programs for simulation of etching and deposition in three dimensions are available from commercial [17] and some academic sources, e.g., [21]–[23]. For the treatment of variations resulting from lithography, etching, or deposition steps, such tools must be employed and linked with equipment simulation to be able to trace the effect of variations, which are caused by process equipment, on device geometries. This is subject to current research and various publications, e.g., [11] and [22].

D. Device Simulation

Most work on semiconductor device simulation currently carried out deals with the development of new or improved transport models and the application of device simulation tools to optimize device architecture and investigate new device options. A key limitation is that the drift-diffusion or hydrodynamic models used in commercial programs such as *Sentaurus Device* [15] or *Victory* [17] fail to match experiments for small feature sizes [24] and therefore need appropriate calibration or extension. More advanced—and, in general, more time consuming—models and device simulators are available from various academic sites, e.g., [25].

Four main aspects are important for the study of variations at the device level. First, the nominal devices must be appropriately described, e.g., using well-calibrated models, as previously mentioned. Second, variations generated during wafer processing need the (mostly 3-D) coupled simulation of the process sequence and the device properties. Third, some variations such as dopant fluctuations need not to be considered in process simulation because their result can be well described by simple statistical distributions, defined only by the number of ions or the absolute value and the correlation length of the LER. In turn, such variations have been studied for many years by device simulation only and are subject to numerous publications, e.g., [3]–[5]. Fourth, temporal variations of the device performance particularly result from aging effects that are, to some extent, also influenced by process variations. A detailed discussion of variability not only extending from device to design but also stressing the importance of process variations has been recently published by Purdue University and Intel [26].

E. Circuit Simulation

The ongoing shrinkage process in technology development leads to an increasing influence of process variations on circuit characteristics [27], [28]. In order to examine these influences, parametric variations, as they appear from tolerances in the manufacturing process, must be explored on the circuit level [29]. The results of the so-called corner simulations [30] are usually too pessimistic since the probability of their occurrence is extremely low. Moreover, the corner case of the process parameters does not necessarily cover the worst case of the circuit performance.

Most commercial SPICE-like circuit simulators offer analysis modes for parametric variations by means of Monte Carlo simulation. Unfortunately, only simple standard distributions (uniform, normal, and log-normal) are provided to describe the probability density function of the random variables. Moreover, Monte Carlo simulations are very demanding (in terms of time and/or computational costs) for complex designs and multidimensional parametric dependences. Even in the presence of FastSPICE simulators [31], which offer a speed-up of one order of magnitude on the average compared with standard SPICE simulators, the conceptual problem persists.

Therefore, efficient analysis methods are required to examine the statistical behavior of circuit characteristics and their dependences on process variations [32]. Methods that avoid expensive Monte Carlo simulations, such as importance sampling and working with marginal distributions, have to be considered for circuit- and system-level analyses. Another approach is to use statistical modeling methods for the description of the dependences over multiple levels of abstraction, such as response surface models with higher order sensitivities. Models of higher order are necessary to sufficiently describe the nonlinearities in the statistical dependences, which come from the involved device models [33]. The nonlinearities lead to distorted probability distributions, as compared with the original normal distributions of the process variables. Therefore, description and analysis methods for strongly asymmetrical parameter variations are also required [34].

F. Mixed-Level Simulation

Classical circuit simulation (sometimes also called SPICE level) uses device models in form of pseudocircuits to map the electrical behavior of each single semiconductor device into the world of netlists. Such a method is limited in its accuracy by nature. Moreover, it requires some effort to generate these models and calibrate their parameters that are up to several hundreds for a particular semiconductor device. In order to obtain stable and/or more accurate results, it is often necessary to insert the system of partial differential equations describing the electrical behavior directly into the set of circuit equations (differential algebraic equations), resulting in a system of tightly coupled partial differential algebraic equations. This approach has been implemented in the simulation framework *Multiphysical Electric Circuit Simulator (MECS)* [35].

MECS has the ability to include device models from several existing simulators, such as SPICE (e.g., *BSIM3* [36]), *MinimosNT* [37] (for the devices to be accurately modeled), or *Magwel* ^{RT} (for electromagnetic effects) [38]. The actual implementation was not carried out by copying or reimplementing parts of the code but by using the aforementioned simulators in form of shared libraries. Since *MECS* is written in Python/Cython, there are easy-to-use mechanisms to import functions from shared libraries originally generated from, e.g., C++ or Fortran. This is not only a practical way of reusing the code but also a powerful method to speed up a Python program.

The resulting series of large nonlinear systems has to be linearized and finally solved by an efficient sparse matrix solver. The *SAMG* software library of Fraunhofer SCAI [39] has been extended by means of a switching and smoothing framework (α-*SAMG* [40], [41]) and physically oriented reformulations in order to speed up the most time-consuming part of the simulation runs.

G. Behavioral Modeling for System Simulation

Growing standards in the design process of electronic circuits require new methods that can improve design efficiency. The functional verification by simulation is an important substep in checking the specification consistency of designed circuit components. On the transistor level, this is associated mostly with very long simulation times. Thus, in larger designs, time-critical components are replaced by manually generated behavioral models. Thereby, the physics of a component is mapped to mathematical equations that describe its system behavior. This leads to improved insights regarding the behavior of the modeled component and an advanced analysis of its influences on the circuit. Due to difficulties in the automated generation of behavioral models, these have to be manually created at the moment [42]–[44].

Since, in most cases, generalized modeling languages, such as Verilog-A or very-high-speed integrated circuit hardware description language (VHDL), are used, there exists a large number of simulators, such as *Spectre*, *Saber*, or *Eldo*, which can integrate and evaluate these models. However, most simulators, which are optimized to handle netlist-based input, tend to handle behavioral models far too slowly [45].

Some simulators allow the use of statistically distributed model parameters instead of just nominal model parameters. These can be used for statistical analyses, e.g., for Monte Carlo simulations. In particular, the software tool *Analog Insydes* [46], which is designed for symbolic modeling and analysis of electronic circuits, unifies these features. Due to the symbolic modeling concept, it is possible to analyze the exact behavior of the circuit. Numerical as well as symbolic methods are used to determine the effects of statistically disturbed model parameters. Furthermore, *Analog Insydes* offers the possibility to reduce behavioral models. Thereby, only the most relevant parts of the model are kept so that the order of the model decreases. This leads to faster simulation and a simplified analysis that allows producing interpretable design formulas from the reduced model.

Currently, the resulting deviation in the output behavior of the nominal system is controlled during model reduction, such that no statements about the behavior of the circuit with disturbed parameters can be made [47].

H. Integration of Simulation Levels

For the investigation of the impact of process variability on circuits and systems, a complete simulation sequence is needed, which spans from equipment and lithography simulation through the overall process and device simulation to compact and behavioral models. Commercial software packages currently cover most of this wide area to some extent but nevertheless have serious limitations in two respects: First, as previously outlined, there are already considerable limitations

$$R^0 \xrightarrow[P^1]{} R^1 \xrightarrow[P^2]{} R^2 \ldots \xrightarrow[P^n]{} R^n \xrightarrow[DE]{} R^{n+1}$$

P^k: Processing step k
DE: Device simulation + SPICE parameter extraction

$$R^k = \begin{Bmatrix} Geometry \\ Doping \end{Bmatrix} \qquad R^{n+1} = \begin{Bmatrix} SPICE - \\ Parameters \end{Bmatrix}$$

Fig. 3. Notation for simulation flow without variations up to the SPICE level.

in the individual simulation steps. Second and more important in this context, the integration with equipment simulation is largely missing, and the integration with lithography and topography simulation has severe limitations. For instance, in commercial software systems, either lithography [17] or overall topography steps [15] can be only simulated in 2-D.

IV. Hierarchical Simulation of Variability

As previously outlined, variability must be traced through all processing steps and simulation levels, from equipment to devices, circuits, and systems. A key additional requirement to be met is that it must be also possible to trace correlations, e.g., if different transistors are patterned with the same lithography step (but different features on the mask) and are therefore subject to the same focus and dose variations. A simple example is the matching of two nominally identical transistors. If both are affected by the same variations, no matching problem occurs. This also demonstrates that the corner analysis can lead to too pessimistic results.

In the following, a novel approach is presented, which allows tracing not only variability but also the respective correlations through all steps in the simulation sequence, from process up to circuit simulation. First, the nominal process and simulation flow is considered without variations (see Fig. 3). Here, the result of the simulation step k is described by a state vector R^k, which contains all relevant information. The simulation step $k+1$ then transforms the state vector R^k into the next one, i.e., R^{k+1}. The content of the state vectors is specific, e.g., for the result of a processing step, it may be the reference to a file, where the resulting geometry and doping concentrations are stored, possibly including also values such as extracted CDs. For the result of a device simulation, it may be the output characteristics of the device and/or extracted compact model parameters (see Fig. 3). Since the content of R^k is freely defined to appropriately characterize the result of the equipment, process or device, this approach is completely versatile and not subject to limitations.

To proceed from device to circuit simulation, the state vectors R^k must be available for all elements of the circuit. Depending on the nature of the circuit element and the way it was fabricated, these vectors may differ even in format (e.g., between a transistor, a resistor, or an interconnect structure). The input needed for the circuit model is then taken from the respective vectors R^k. This allows differentiating between statistically independent and dependent variations: If two circuit elements are not only nominally identical but have been also subject to identical variations, then the same variation of the state vector R^k has to be used for both elements. Otherwise, two different representatives of R^k must be used, both describing the same device but considering different variations. Whether the variations of different circuit elements are statistically independent or not depends on the layout and the details of the process sequence. Basically, variations that have their source in different processing steps (e.g., lithography and implantation, or two lithography steps for two different mask levels, which, in turn, may have different focus and dose values) are statistically independent from each other. Therefore, it is essential for treating variations at the circuit level to know if the different circuit elements have variation-critical steps in common and therefore have correlated variations. The generalization to variations that have some correlation length is obvious but, for simplicity, not further explained here.

$$\begin{aligned} R^0 \xrightarrow[P^1]{} R^1 \xrightarrow[P^2]{} R_0^2 \ldots \xrightarrow[P^n]{} R_0^n \xrightarrow[DE]{} R_0^{n+1} \\ \ddots \quad \ddots R_1^2 \ldots \xrightarrow[P^n]{} R_1^n \xrightarrow[DE]{} R_1^{n+1} \\ \ddots R_2^2 \ldots \xrightarrow[P^n]{} R_2^n \xrightarrow[DE]{} R_2^{n+1} \end{aligned}$$

P^k: Process step k
DE: Device simulation + SPICE parameter extraction
$j^{th}\,\Delta p$: j^{th} value for the variation of one process parameter p

$$R_j^2 = \begin{Bmatrix} j^{th}\Delta p \\ Geometry \\ Doping \end{Bmatrix} \qquad R_j^{n+1} = \begin{Bmatrix} j^{th}\Delta p \\ SPICE - \\ Parameters \end{Bmatrix}$$

Fig. 4. Notation for the simulation flow with variation of processing step 2 up to the SPICE level. Three different values of one parameter are shown.

To trace variations along the whole simulation sequence, first those process parameters have to be identified for which variations have to be considered, e.g., the defocus in a critical lithography step or the variations of temperatures of a source/drain annealing step. The simulation sequence is then split at the first occurrence of a variation to be treated, with an appropriate assumption on the statistics of a varying parameter (e.g., focus and dose of a lithography steps both independently varying and homogeneously distributed with a given interval). This is visualized in Fig. 4 for the second processing step P^2, which is simulated with (cf. Fig. 3) different values of one parameter p, resulting in (here three) different state vectors R_j^2 after the simulation step. These vectors R_j^2 all refer to the same nominal simulation sequence but with different variations. Obviously, the value of the varying parameter should be then also stored in vector R_j^2. In turn, at this step the simulation is split into as many individual simulations as the parameter values used to characterize the variation.

During a full hierarchical simulation sequence, each new variation considered leads to an additional split of the simulation sequence. If, for example, five varying parameters are considered, each of them with 20 values, this finally leads to 3.2 million vectors after the last simulation step. Obviously, the treatment of a variation is more costly the earlier it turns up in the simulation sequence.

The final set of vectors R^k after the last simulation step then allows for the extraction of all desired statistics: The overall distribution of a device parameter (e.g., $V_{\rm th}$) or all sets of multidimensional or conditioned distributions (e.g., distribution of $I_{\rm on}$ for all devices that have $I_{\rm off}$ lower than a given limit).

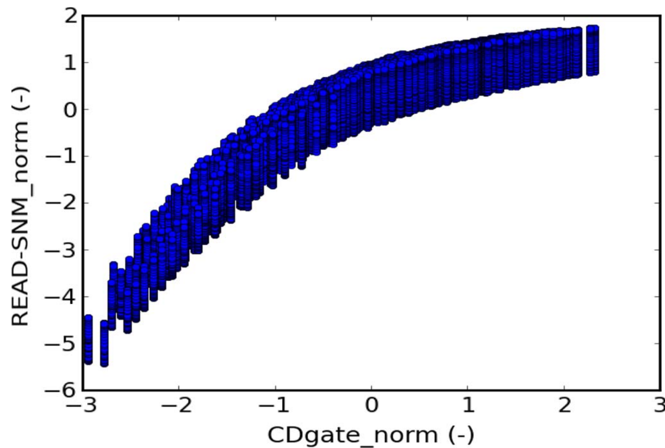

Fig. 5. Example for the impact of six process variations studied in [2] on the READ static noise margin of a six-transistor SRAM cell with a 32-nm nominal-transistor gate length. Normalization explained in the text below.

Obviously, it is very advantageous to reduce the complexity as soon as possible along the simulation sequence, e.g., by deriving variation-aware compact models from device simulation [48]. In particular, an appropriate reduction of the full (5-D here) matrix of simulations, e.g., by design of experiment (DoE) methods, is very promising to drastically reduce computation times. As an enhancement of the split, *PRO-CHAIN* [49], [50] can be adopted (see [2, Sec. III]).

As a simple approach, linearization may be also beneficial to reduce the number of simulations if different sources of variation can be independently treated. In that case, not the whole matrix of combinations of variations needs to be considered, but each variation could be separately treated, and its impact could be approximately described by two or a few simulations only. If, however, nonlinearities or correlations dominate the block behavior, the use of DoE methods as previously mentioned is necessary to speed up the study.

Fig. 5 gives a brief example of results obtained with this approach, where the impact of variations of dose and focus for two lithography steps, of the etch bias across the wafer, and of the peak annealing temperature on a static random-access memory (SRAM) cell were considered. The figure shows the normalized value of the READ static noise margin versus the normalized variations of the gate length. Here, "normalized" refers to the deviation of the actual value from its average, divided by its standard deviation. As visualized in the figure, the READ static noise margin strongly depends on variations of the gate length. However, since different values of the six process variations considered lead to the same value of the gate length and particularly because the peak annealing temperature is varied as well, a spread of the READ static noise margin for the constant gate length is also observed. For more explanations and results, see [2].

V. Conclusion

A large diversity of variabilities challenges the manufacturability of advanced nanoelectronic devices, circuits, and systems. For current bulk transistors, dopant fluctuations are the dominant source of variations, which have been broadly discussed in the literature. However, many variations caused by fabrication equipment have to be taken into consideration, particularly for upcoming SOI transistors with undoped channels. Hierarchical coupled equipment, process, device, circuit, and system simulation is a valuable approach to assess the impact of variations and to decide about the best equipment, process, and device architecture options and the best tradeoffs to be taken. Examples for simulation results obtained at Fraunhofer are shown in the consecutive paper presented in the same issue of this journal.

Acknowledgment

The authors would like to thank their colleagues at Fraunhofer A. Burenkov, C. Kampen (IISB), C. Sohrmann (IIS/EAS), U. Paschen (IMS), M. Hauser (ITWM), B. Klaassen (SCAI), and C. Tischendorf (Univ. Köln) for their contributions.

References

[1] The International Technology Roadmap on Semiconductors, 2009 issue. [Online]. Available: www.itrs.net
[2] J. Lorenz, E. Bär, T. Clees, P. Evanschitzky, R. Jancke, C. Kampen, U. Paschen, C. Salzig, and S. Selberherr, "Hierarchical simulation of process variations and their impact on circuits and systems: Results," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2227–2234, Aug. 2011.
[3] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, "Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs," *IEEE Trans. Electron Devices*, vol. 50, no. 9, pp. 1837–1852, Sep. 2003.
[4] K. Takeuchi, A. Nishida, and T. Hiramoto, "Random fluctuations in scaled MOS devices," in *Proc. SISPAD*, San Diego, CA, 2009, pp. 79–85.
[5] D. Reid, C. Millar, G. Roy, S. Roy, and A. Asenov, "Understanding LER-induced statistical variability: A 35,000 sample 3D simulation study," in *Proc. ESSDERC*, Athens, Greece, Sep. 14–18, 2009, pp. 423–426.
[6] C. Mazure, R. Ferrant, B.-Y. Nguyen, W. Schwarzenbach, and C. Moulin, "FDSOI: From materials to devices and circuit applications," in *Proc. ESSDERC*, Sevilla, Spain, 2010, pp. 45–51.
[7] C. Kampen, A. Burenkov, J. Lorenz, and H. Ryssel, "FDSOI MOSFET compact modeling including process variations," in *Proc. ULIS Conf.*, Glasgow, U.K., 2010, pp. 173–176.
[8] T. Schnattinger, E. Bär, and A. Erdmann, "Mesoscopic resist processing in optical lithography," in *Proc. SISPAD*, Monterey, CA, 2006, pp. 341–344.
[9] P. Pichler, *Intrinsic Point Defects, Impurities, and Their Diffusion in Silicon*. New York: Springer-Verlag, 2004.
[10] ESI Group, Simulator CFD-ACE+2009.
[11] E. Bär, D. Kunder, P. Evanschitzky, and J. Lorenz, "Coupling of equipment simulation and feature-scale profile simulation for dry-etching of polysilicon gate lines," in *Proc. SISPAD*, Bologna, Italy, 2010, pp. 57–60.
[12] S. K. Jones and C. Hill, "Modeling of temperature non-homogeneity during rapid thermal engineering," in *Proc. UK Alvey Process Device Modelling Club Meeting*, Swansea, U.K., 1988.
[13] [Online]. Available: www.kla-tencor.com
[14] [Online]. Available: www.panoramictech.com
[15] [Online]. Available: www.synopsys.com
[16] P. Evanschitzky, A. Erdmann, and T. Fühner, "Extended Abbe approach for fast and accurate lithography imaging simulations," in *Proc. SPIE*, Dresden, Germany, 2009, 7470, pp. 1–11.
[17] [Online]. Available: www.silvaco.com
[18] [Online]. Available: www.drlitho.com
[19] P. Evanschitzky, F. Shao, A. Erdmann, and D. Reibold, "Simulation of larger mask areas using the waveguide method with fast decomposition technique," in *Proc. SPIE*, Grenoble, France, 2007, vol. 6730, pp. 1–9.
[20] [Online]. Available: www.atemox.eu
[21] E. Bär, J. Lorenz, and H. Ryssel, "Three-dimensional simulation of layer deposition," *Microelectron. J.*, vol. 29, no. 11, pp. 799–804, Nov. 1998.
[22] A. Labun, H. Moffat, and T. Cale, "Mechanistic feature-scale profile simulation of SiO_2 low-pressure chemical vapor deposition by

tetraethoxysilane pyrolysis," *J. Vac. Sci. Technol. B, Microelectron. Nanometer Struct.*, vol. 18, no. 1, pp. 267–278, Jan. 2000.
[23] O. Ertl and S. Selberherr, "Three-dimensional level set based bosch process simulations using ray tracing for flux calculation," *Microelectron. Eng.*, vol. 87, no. 1, pp. 20–29, Jan. 2010.
[24] T. Grasser, C. Jungemann, H. Kosina, B. Meinerzhagen, and S. Selberherr, "Advanced transport models for sub-micrometer devices," in *Proc. SISPAD*, München, Germany, 2004, pp. 1–8.
[25] [Online]. Available: http://nanohub.org/
[26] S. Ghosh and K. Roy, "Parametric variation tolerance and error resiliency: New design paradigm for the nanoscale era," *Proc. IEEE*, vol. 98, no. 10, pp. 1718–1751, Oct. 2010.
[27] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variations and impact on circuits and microarchitecture," in *Proc. DAC*, Anaheim, CA, 2003, pp. 338–342.
[28] S. K. Springer, S. Lee, N. Lu, E. J. Nowak, J.-O. Plouchart, J. S. Watts, R. Q. Williams, and N. Zamdmer, "Modeling of variation in submicrometer CMOS ULSI technologies," *IEEE Trans. Electron Devices*, vol. 53, no. 9, pp. 2168–2178, Sep. 2006.
[29] U. Eichler, J. Haase, R. Häuler, and H. Kinzelbach, "Gate-level digital power simulation with varying technology parameters," in *Proc. SCD*, Dresden, Germany, 2008.
[30] J. Salzmann, F. Sill, and D. Timmermann, "Algorithm for fast statistical timing analysis," in *Proc. Int. Symp. SoC*, Tampere, Finland, 2007, pp. 1–4.
[31] S. Sangameswaran and S. Yamauchi, "Post-layout parasitic verification methodology for mixed-signal designs using fast-SPICE simulators," in *Proc. IEEE Dallas/CAS Workshop*, 2005, pp. 211–214.
[32] A. Srivastava, D. Sylvester, and D. Blaauw, *Statistical Analysis and Optimization for VLSI: Timing and Power*. New York: Springer-Verlag, 2005.
[33] C. Sohrmann, L. Muche, and J. Haase, "Accurate approximation to the probability of critical performance," in *Proc. 2nd GMM/GI/ITG—Fachtagung Zuverlässigkeit und Entwurf*, Ingolstadt, Germany, 2008, pp. 93–97.
[34] M. Zhang, M. Olbrich, D. Seider, M. Fredrichs, H. Kinzelbach, and E. Barke, "Ein Verfahren zur Analyse von Prozessschwankungen für nichtlineare Schaltungen mit nicht-Gauss-verteilten Parametern," in *Proc. 1st GMM/GI/ITG—Fachtagung Zuverlässigkeit und Entwurf*, München, Germany, 2007, pp. 25–30.
[35] M. Selva Soto and C. Tischendorf, "Numerical analysis of DAEs from coupled circuit and semiconductor simulation," *Appl. Numer. Math.*, vol. 53, no. 2–4, pp. 471–488, May 2005.
[36] [Online]. Available: www-device.eecs.berkeley.edu/~bsim3/get.html
[37] [Online]. Available: www.iue.tuwien.ac.at
[38] [Online]. Available: www.magwel.com
[39] K. Stüben and T. Clees, *SAMG User's Manual v. 22c*. Sankt Augustin, Germany: Fraunhofer Inst. SCAI, 2005.
[40] T. Clees, T. Samrowski, M. Zitzmann, and R. Weigel, "An automatic multi-level solver switching strategy for PEEC-based EMC simulation," in *Proc. IEEE 18th Int. EMC-Zurich*, München, Germany, 2007, pp. 25–28.
[41] P. Thum, T. Clees, G. Weyns, G. Nelissen, J. Deconinck, "Efficient algebraic multigrid for migration-diffusion-convection-reaction systems arising in electrochemical simulations," *J. Comput. Phys.*, vol. 229, no. 19, pp. 7260–7276, Sep. 2010. [Online]. Available: http://dx.doi.org/10.1016/j.jcp.2010.06.011
[42] L. Gomez and J. M. Fernandes, Eds., *Behavioral Modeling for Embedded Systems and Technologies: Applications for Design and Implementation*. Hershey, PA: IGI Global, 2010.
[43] R. Sommer, T. Halfmann, and J. Broz, "Automated behavioral modeling and analytical model-order reduction by application of symbolic circuit analysis for multi-physical systems," *Simul. Model. Pract. Theory*, vol. 16, no. 8, pp. 1024–1039, Sep. 2008.
[44] R. Narayanan, N. Abbasi, M. Zaki, G. Al Sammane, and S. Tahar, "On the simulation performance of contemporary AMS hardware description languages," in *Proc. ICM*, Sharjah, UAE, 2008, pp. 361–364.
[45] F. M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE Microw. Mag.*, vol. 10, no. 7, pp. 52–64, Dec. 2009.
[46] [Online]. Available: www.analog-insydes.de
[47] T. Halfmann, J. Broz, C. Knoth, D. Platte, and P. Rotter, "Generation of efficient behavioral models using model compilation and model reduction techniques," in *Proc. 10th Int. Workshop SMACD*, Erfurt, Germany, 2008, pp. 203–209.
[48] D. Steffes-lai, L. Nikitina, and T. Clees, "Vorrichtung und Verfahren zum Bearbeiten einer Prozess-Simulationsdatenbasis eines Prozesses," PCT/EP2010/061450, Aug. 5, 2010.
[49] D. Steffes-lai and T. Clees, "Efficient stochastic analysis of process chains," in *Proc. 1st Int. Fraunhofer Multiphys. Conf.*, Bonn, Germany, 2010, *Int. J. Multiphys. Simul.*, accepted for publication.
[50] R. Jancke, C. Kampen, O. Kilic, and J. Lorenz, "Hierarchischer Ansatz für die Monte-Carlo-Simulation komplexer Mixed-Signal-Schaltungen," in *Proc. 11th ITG/GMM—Fachtagung ANALOG*, Erfurt, Germany, 2010, pp. 185–190.

Jürgen K. Lorenz (M'01) was born in Stockelsdorf, Germany, in 1957. He received the "Diplom-Mathematiker" and "Diplom-Physiker" degrees from the Technical University of Munich, Munich, Germany, in 1982 and 1984, respectively, and the Dr.-Ing. degree in electrical engineering from the University of Erlangen–Nuremberg, Erlangen–Nuremberg, Germany, in 2000.

He joined Fraunhofer in 1983. Since 1985 he is in charge of the technology simulation department of the then newly founded Fraunhofer IISB in Erlangen. His main subjects are the development of physical models and programs for semiconductor process simulation and the required algorithms. He authored or co-authored about 120 papers. During the last 18 years he has been involved in 27 European projects, for 7 of which he acted as coordinator.

Dr. Lorenz is member of the Electrochemical Society, and has repeatedly been or is a member of the technical committees of the ESSDERC, SISPAD, and IEDM conferences. Following requests from industry, he has been contributing since 2000 as expert to the preparation of the International Technology Roadmap on Semiconductors and has been the chairman for its Modeling and Simulation chapter since 2002.

Eberhard Bär was born in Darmstadt, Germany, in 1967. He received the "Diplom-Physiker" degree with specialization in experimental solid-state physics from Darmstadt University of Technology, Darmstadt, in 1992 and the Dr.-Ing. degree in electrical engineering with specialization in semiconductor technology from the University of Erlangen–Nuremberg, Erlangen–Nuremberg, Germany, in 1998.

Since 1992, he has been with the Fraunhofer Institute for Integrated Systems and Device Technology (IISB), where he is currently in charge of the topography simulation group. His current research interests focus on the development and application of topography simulation tools in semiconductor technology. He is author or co-author of 40 publications in journals and conference proceedings.

Tanja Clees was born in Aachen, Germany, in 1977. She received the "Diplom-Mathematikerin" degree with minor subject of physical chemistry and the Dr. rer. nat. degree in applied mathematics from the University of Cologne, Cologne, Germany, in 1999 and 2004, respectively.

She joined Fraunhofer SCAI (till 2001 GMD SCAI) in Sankt Augustin, Germany in 1998. From 2000 to 2006 she worked as a scientist for and chief software developer of the linear solver library *SAMG*. Since 2007, she has been in charge of the robust design group and its two software packages *DesParO and DiffCrash*. Her main subjects are hierarchical solvers for large sparse linear systems, as well as statistical analysis and exploration of dependences of highly resolved simulation results from parameter variations and robust multiobjective parameter optimization.

Roland Jancke was born in Berlin, Germany, in 1970. He received the Diploma in electrical engineering from the Technical University of Dresden, Dresden, Germany, in 1996.

Since 1996, he has been with the Division Design Automation of the Fraunhofer Institute for Integrated Circuits (IIS/EAS), Dresden, where he is currently heading the group for technology oriented modeling. His main scientific interest is in behavioral modeling of analog and mixed-signal circuits, with focus on reliability and technology-induced effects.

Mr. Jancke is member of the Verband der Elektrotechnik, Elektronik und Informationstechnik and actively participating in its working group on design methods for analog circuits.

Christian P. J. Salzig was born in Boppard, Germany, in 1980. He received the Diplom-Technomathematiker degree with minor subject of electrical engineering from the Technische Universität (TU) Kaiserslautern, Kaiserslautern, Germany.

From 2005 to 2008, he was awarded with a Ph.D. scholarship of the Graduate School with the Department of Mathematics, TU Kaiserslautern. Since 2008, he has been a scientific assistant at the Fraunhofer Institute for Industrial Mathematics (ITWM). His main subjects are statistics, modeling of nonlinear systems, and control theory.

Mr. Salzig has been a member of the developer team of the symbolic electronic design automation tool *Analog Insydes* since 2001.

Siegfried Selberherr (M’79–SM’84–F’93) was born in Klosterneuburg, Austria, in 1955. He received the Diplomingenieur degree in electrical engineering and the Ph.D. degree in technical sciences from the Technische Universität Wien, Vienna, Austria, in 1978 and 1981, respectively.

Since 1984, he has been holding the *venia docendi* on computer-aided design. Since 1988, he has been the Chair Professor of the Institut für Mikroelektronik. From 1998 to 2005, he was the Dean of the Fakultät für Elektrotechnik und Informationstechnik. His current research interests are modeling and simulation of problems for microelectronics engineering.

Dr. Selberherr is a member of the ACM, the ECS, the SIAM, and the VDE.