Modelling and Analysis of Manufacturing Variability Effects from Process to Architectural Level

Chenxi Ni, Ziyad Al Tarawneh, Gordon Russell, and Alex Bystrov

MSD Group, School of EEE, Newcastle University, Newcastle upon Tyne, UK
{Chenxi.ni,Ziyad.al-tarawneh,G.Russell,A.Bystrov}@ncl.ac.uk

Abstract. This paper describes the development of a cell library which can be used to efficiently predict the distribution of circuit delay and leakage power performance due to process variation effects. In developing the library a step-wise approach is adopted in which the effects of process variations on the design parameters of interest at the various levels of design abstraction are evaluated, that is from transistor through circuit to architectural level. A cell library is generated comprising functional blocks whose complexity ranges from a single gate up to several thousand gates. As a demonstration vehicle a 2-stage asynchronous micropipeline is simulated using the cell library to predict the subsequent delay and leakage power distributions. The experimental results show that the proposed method is much faster than the traditional statistical static delay/power analysis (SSTA/SPA) approaches by a factor of 50; the results are also compared with Monte Carlo simulation data for validation purposes, and show an acceptable error rate of within 5%.

1 Introduction

When there is a degree of uncertainty in the fabrication process, the potential effects of the fluctuations implicitly impact not only on device manufacturability and reliability but also on design 'aggressiveness' which affects design performance and subsequently the profitability of the final product.

The statistical variation analysis from process parameters to device performance is commonly based on the Design of Experiment (DoE) technique and Response Surface Methodology (RSM) [2], which enables designers to construct empirical models from which the output responses can be determined as a function of the input factors or parameters. The variability effects from device level to circuit level can be analyzed using Statistical Static Timing Analysis (SSTA) and Statistical Power Analysis (SPA) techniques [3-10], wherein the variation sources are described as Random Variables (RVs) and the performance parameters are usually modelled as low-order polynomials of all the RVs [3].

Although a lot of research into process variation effects has been made at each stage of analysis, there are very few methods which provide a complete methodology to model these effects from the perspective of a designer in terms of circuit performance factors. On the other hand, at present, most of the investigations into statistical

J.L. Ayala, D. Shang, and A. Yakovlev (Eds.): PATMOS 2012, LNCS 7606, pp. 11–20, 2013.
© Springer-Verlag Berlin Heidelberg 2013

delay and leakage power models have been limited to small circuit models. In order to perform an efficient variation analysis on a large circuit, it is essential to take into consideration the effects of process variation on delay and leakage power at a higher level of design abstraction.

This paper introduces a statistical approach based on Technology Computer Aided Design (TCAD) tools and statistical techniques to model the impact of process variation effects on the device performance metrics for devices realised by bulk Silicon process technology. This technique has been applied to the modelling of the variation effects from 16 typical process parameters on the zero-biased threshold voltage, V_{th0}, of the devices. The accuracy of this technique is checked using the 'goodness' of the second-order fit, and the experimental results show that the errors of the mean and sigma value of V_{th0} estimated using the proposed methodology are within 4%. Furthermore, a statistical cell library characterization methodology is proposed that efficiently migrates the effect of process variation on delay and leakage power at device and circuit level to higher levels of abstraction where the overall effect on system performance can be analyzed and design modifications made to ameliorate these effects early in the design cycle. As a demonstration vehicle, the models have been implemented in a 2-stage pipeline circuit. The experimental results show that this methodology can achieve a relatively high accuracy in which the error in the mean and sigma values for delay and leakage power are under 5% compared with 5000-sample Monte Carlo (MC) data, with a computation time which is at least 50 times faster than traditional SSTA/SPA approaches.

2 Analysis of Process Variation Effects on Device Performance

A statistical approach based on TCAD and statistical techniques to model the impact of process variation effects on the device performance metrics for NMOS and PMOS transistors realised by bulk Silicon process technology, is outlined in this section. The general methodology for studying variability is shown in Fig.1 and involves parameter screening, model building and model analysis. The methodology begins with the calibration of the TCAD process and device electrical characteristics with the experimental data, and the extraction of the compact model parameters for nominal devices. During the process simulation, to generate an accurate set of device characteristics, all necessary physical models were incorporated in order to have as realistic a simulation as possible. The basic model for the complete process simulation consisted of a diffusion model with charged point defects, a transient dopant clustering model, a three-phase segregation model and a mechanical stress model, including the thermal and lattice mismatch as well as intrinsic stresses. Thereafter the compact models for NMOS and PMOS transistors were extracted for the nominal devices using I-V data generated.

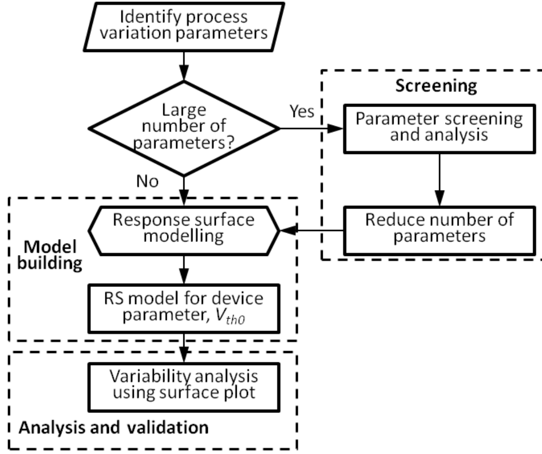

Fig. 1. Flow chart of variability analysis utilized by DoE and RSM statistical techniques

The compact model parameter chosen in this work is V_{th0} which represents the threshold voltage for long channel devices at zero bias voltage. The reason for choosing V_{th0} out of the numerous other compact model parameters is that it shows a strong statistical relationship with circuit performance metrics.

3 Analysis of Process Variation Effects on Circuit Performance

In this section, a description is given to the first-order statistical modelling methodology, which will be applied to model the process variation effects on circuit performance in terms of propagation delay time and leakage power dissipation, followed by a general description of the corresponding statistical analysis techniques.

3.1 Statistical Gate Delay and Leakage Power Models

In statistical gate performance modelling, device and circuit environmental parameters will be represented by a Random Variables (RV), which are usually assumed to be Gaussian. The circuit performance parameters are modeled as low-order polynomials of the source RVs. Equations (1) and (2) show the 1st order canonical form for the delay and leakage power models [4][10].

$$D(Delay) = \mu_D + \sum_{i=1}^{n} \beta_{Di} G_i + \beta_{D(n+1)} R \quad (1)$$

$$LP(Leakage\ Power) = exp\left(\mu_P + \sum_{i=1}^{n} \beta_{Pi} G_i + \beta_{P(n+1)} R\right) \quad (2)$$

Equation (1) is the first-order canonical gate delay model. μ_D is the mean delay time of the gate. G_i represents the ith global variational source (Inter-die). R is the sum of all the local RVs in the gate (Intra-die). $\beta_D s$ are the sensitivity coefficients for all the

RVs in this delay model. The gate leakage power on the other hand is modelled as a lognormal RV in Equation (2) since the gate leakage current has an exponential relationship with the variational sources. All the RVs in Equations (1) and (2) follow a normal distribution (Gaussian).

3.2 Statistical Timing and Power Analysis

Both of the timing and leakage power analysis techniques can be used to estimate the overall probability density functions (PDFs) of multiple cell models in terms of delay and power. In order to keep the delay and power analysis alive, both of the techniques use a corresponding approximation methodology to define the non-normal analysis result in a normal canonical form. A small error will be introduced by doing so; however this is not significant (within 5% as shown in the experimental results in section 5). The tightness probability based SSTA approach from C.Visweswariah [4] and the recursive moments-matching based SPA technique from A.Srivastava et al [9] are employed in this work. Having established the cell model form and the analysis methods, the device parameter variation effects can be analyzed on circuit performance in terms of delay and leakage power dissipation. Furthermore, a statistical cell library for 90nm technology has been built to bring the process variation effects up to the architectural level.

4 Analysis of Process Variation Effect at Architectural Level

In order to model process variation effects at architectural level, a statistical cell library comprising a variety of functional blocks has been built. In this section, it will discuss the characterization of the library cells will be discussed in detail.

4.1 Cell Characterization

The gate delay is not only relevant to the device parameters under variation, but also is highly correlated to its operating conditions such as load capacitance C_L and input signal slope T_{in}. Fig.2 shows this relationship between propagation delay and operating conditions for an inverter circuit. It is very difficult to model the operating condition effects (T_{in} and C_L) on propagation delay in canonical form, typically the table look-up approach will solve this problem, where the delay time is sampled with respect to a range of C_L and T_{in} values, then saved in memory. In this work, 7 typical C_L and T_{in} values are sampled as break points, any delay values in close proximity to these will be estimated from the linear function of its adjacent break points. Fig.3 shows a piecewise linear fitting of the inverter delay which is a more efficient view in terms of volume of data required for its representation.

Unlike the statistical delay cells, the characterization of leakage power does not require to be the modelling of the input signal slope because there is no input signal transition in a static environment; furthermore, the leakage power dissipation of a cell has no relationship with its load capacitance, hence there is no need for it to be

considered. The leakage power of a cell will only differ when the cell's state has been changed. Hence, only one 2-D lookup table is sufficient to characterize the leakage power of a whole cell where the rows represent the different input signal states, and the columns represent the coefficients in the canonical power polynomial form.

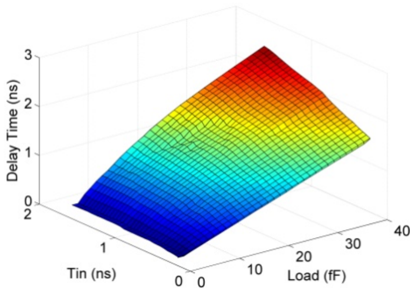

Fig. 2. Inverter delay vs. C_L and T_{in}

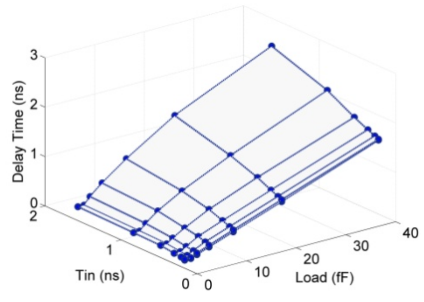

Fig. 3. Modelling using look-up table

4.2 Modelling Higher Level Blocks

By using the methodology introduced in this section, all the basic cells for the logic gates in the library can be created. With the help of the SSTA/SPA techniques, the delay and leakage power performance of any circuit can be analyzed. Consequently, the higher level digital blocks can be modeled using SSTA/SPA analysis results from lower level cells, instead of using SPICE runs; it is very time consuming to run SPICE simulations on larger circuits in order to construct the look-up tables for the delay model. Fig.4 shows a schematic view of variability aware cell modeling framework, which illustrates process variation effects propagating from transistor level to architectural level.

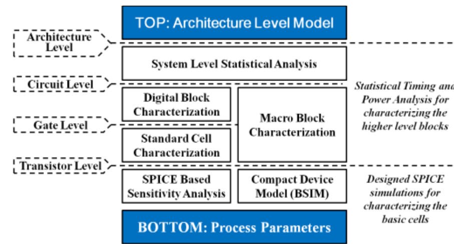

Fig. 4. Schematic view of variability aware cell modeling [12]

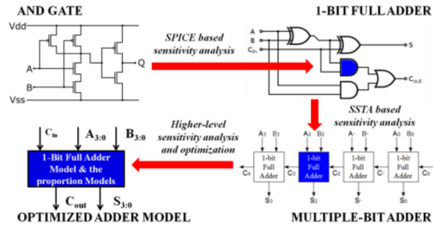

Fig. 5. Characterization Flow from Standard Cells to 4-bit Adder

Once a digital block has been characterized, it can be used as the standard cell to perform SSTA/SPA at a higher level in a more complex circuit, expending the cell library to architectural level blocks in a hierarchical manner. Since only the variability calibrated results of top level digital blocks are used, the models permits a very fast delay analysis to be performed, which also makes it more suitable for scaling up to a larger system. Fig.5 shows an example of the library block characterization flow from

standard cell to a ripple carry adder. The only problem is that a more complex system has a larger number of input transition cases, which leads to a massive memory requirement in order to model each circuit switching case. However, larger functional blocks always have a lot of symmetry and multiple occurrences in the circuit. In Fig.5, the ripple carry adder is actually a serial connection of multiple full adders, so it can be characterized just by the full adder model. During the work of constructing the whole library, most of the blocks can be represented using a smaller circuit model, the output delay time is simply a matter of the proportions of the model.

5 Implementation and Experimental Results

To study and analyze the impact of process variability on the compact model parameter V_{th0}, sixteen process parameters were identified as potential sources of uncontrollable variation at different process steps for the devices implemented in a bulk silicon technology. All the process parameters were varied by ±10% of their mean values. However the process temperatures during the different manufacturing steps were set at ±10 °C from the nominal. This is because the temperature values are very high and, in practice, would not drift in the range of ±10%. The range of variation (± 10% and ±10 °C) corresponds to ±3σ variation. The Pareto plots indicating the relative magnitude of the effects which various process parameters have on V_{th0}, for NMOS and PMOS devices are shown in Fig.6-7.

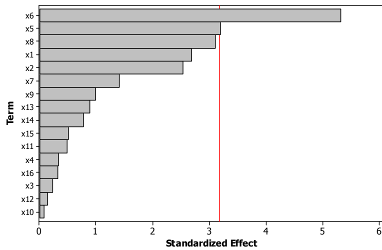

Fig. 6. Pareto plots for NMOS device

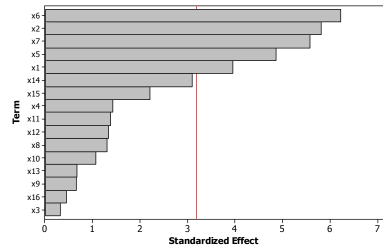

Fig. 7. Pareto plots for PMOS device

Table 1 summarizes the most significant process steps that influence V_{th0} for both NMOS and PMOS devices based on the observation from the Pareto plots.

Table 1. Most significant parameters that impact V_{th0} for NMOS and PMOS devices

NMOS	PMOS
V_{th} adjustment implantation energy (x_6). V_{th} adjustment implantation dose (x_5). Halo implantation energy (x_8). Gate oxide thickness (x_1). High-k dielectric thickness (x_2).	V_{th} adjustment implantation energy (x_6). High-k dielectric thickness (x_2). Halo implantation dose (x_7). V_{th} adjustment implantation dose (x_5).

Having identified the most significant process parameters for the device responses in the screening steps in terms of the compact model parameter, namely V_{th0}, the response surface models for these compact model parameters were subsequently built. Fig.8 and Fig.9 show the response surface for V_{th0} for the NMOS device as a function of different process parameters.

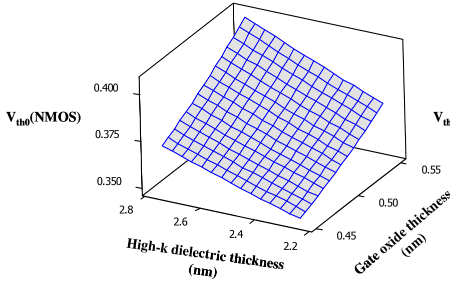

Fig. 8. Response surface for V_{th0} (NMOS) with respect to gate-oxide thickness and high-k dielectric thickness

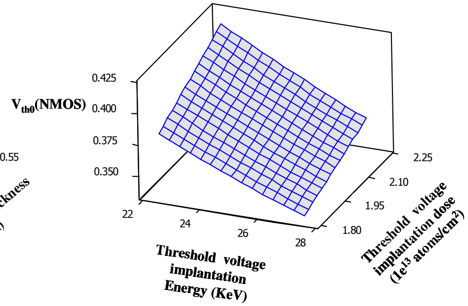

Fig. 9. Response surface for V_{th0} (NMOS) with respect to implantation energy and dose

Table 2. The variation result of V_{th0} for both NMOS and PMOS devices

Device Type	V_{th0}(V)				
	μ	σ	σ /u	R2	R2adj
NMOS	0.3975 v	0.0362 v	10.98%	97.93%	96.05%
PMOS	-0.3957 v	0.049 v	12.38%	98.38%	96.10%

Table 2 compares the μ and σ values of V_{th0} for both NMOS and PMOS devices followed by the 'goodness' of the second-order fit. It can be seen that the *R2* and *R2adj* values are very close to 100% for all responses. This is desirable and therefore ensures that the models accurately highlight the variability due to process fluctuations.

Since the mean and sigma values have been established and verified as accurate, V_{th0} can be assumed as a variation source and applied to the cell models in the statistical cell library which was introduced in Section 4. For demonstration purpose, V_{th0} of the NMOS and PMOS devices are assumed to be normal-distributed local variables with sigma value equal to 11% and 12% of its mean values respectively; additionally the supply voltage V_{dd} and operating temperature T are chosen as the global variation sources, which are also normal variables, which vary ±15% of their mean values. (The mean value of V_{dd} and T is 1v and 75 °C respectively)

Subsequently, the process variation effects were analyzed on a 2-stage asynchronous pipeline circuit using the cell library with the assumed parameter specification above. The pipeline circuit is an event-controlled two-phase bounded data system [11] with an overall complexity of 3272 gates which is shown in Fig.10. It contains 2 asynchronous pipeline registers, a 3-to-8 decoder, a 16-bit register file with 16 memory cells and a

16-bit ALU which can perform 8 logic and arithmetic functions. All the circuit cells were realized using BSIM 4.0 90nm technology model card from University of California, and simulated using Berkeley SPICE program. The cell library is implemented in Matlab Simulink, which provides a convenient graphic interface of all cell blocks to the user. All the transistor parameters in one gate or basic cell are assumed to have a high correlation to each other, and share the same local variables.

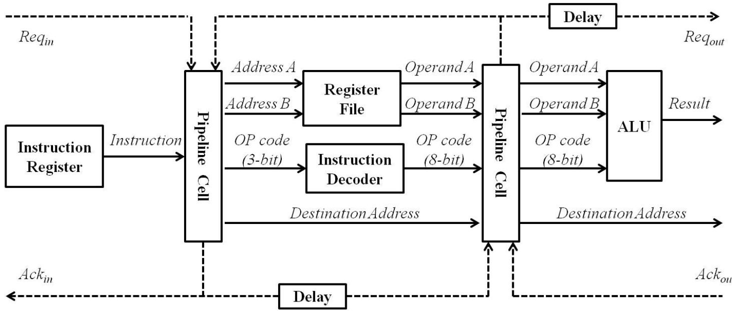

Fig. 10. Pipeline processor block diagram

The pipeline circuit has also been simulated using SPICE-base Monte Carlo technique for validation purposes. As an example, Fig.11 and 12 show the PDFs for the propagation delay time and leakage power dissipation of the ALU block. The histograms in Fig.11 and 12 are generated by a 5000-sampled Monte Carlo simulation, and the solid lines are the predicted PDF obtained using the cell library. It can be seen from the graphs, the predicted PDF fit the Monte Carlo data very well.

All the experimental results are compared with 5000 sampled Monte Carlo simulation, the errors in the delay and leakage power distributions for the main blocks used in the pipeline circuit are listed in Table 3. All the errors are within 5%, 3% for most of the cases. Furthermore, the proposed cell library is also much more feasible compared with MC simulation in analyzing the process variation effects on circuit performance; it would take a month to run a 5000 sampled MC simulation for the pipeline circuit in Fig.10 whereas the cell library only requires a few seconds. The speed-up factors shown in Table 3 have illustrated this advantage of the proposed technique.

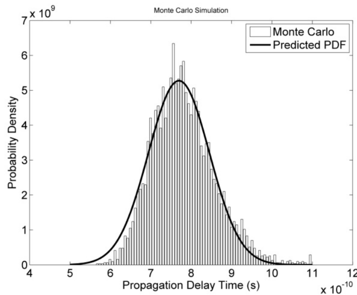

Fig. 11. Delay PDF Matching for ALU

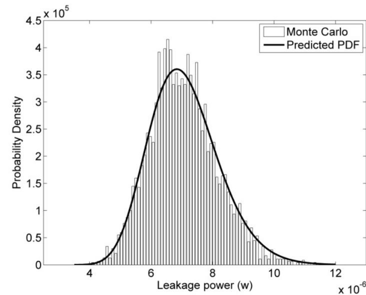

Fig. 12. Leakage PDF matching for ALU

Additionally Table 3 also compares the CPU time between the cell library and traditional SSTA and SPA approaches; it shows that using the cell library is also much faster in computing the delay and leakage power distribution of circuits over traditional SSTA and SPA approaches. The speed-up factor is highly related to the regularity of the circuits. For example, the computation time for the performance analysis of the Register File (RF) block is at least 100 times faster than SSTA and SPA; this is because there are a large number of identical digital blocks (registers) in the circuit of the RF which can be represented by a single model block. On the other hand the speed-up factor for the decoder block is only around 10 since most of the decoder circuit is modeled at gate level. The experimental results show that the overall speed-up factor for the whole demonstration pipeline circuit is more than 50.

Table 3. Comparison of Results

Blocks	**No. of Gates**	**Compared with 5000 Sampled MC Simulations**					**Computation Time Comparison**	
		Delay Error		**Leakage Error**		**Speed-up Factor**		
		μ (%)	**σ (%)**	**μ (%)**	**σ (%)**		**Cell Library**	**SSTA & SPA**
Decoder	*21*	*1.26*	*4.65*	*1.76*	*1.89*	*9,543*	*0.84s*	*11.8 s*
ALU	*743*	*0.62*	*2.32*	*0.73*	*3.45*	*20,246*	*2.40s*	*33.9 s*
Register File	*2574*	*0.79*	*1.90*	*1.87*	*3.49*	*28,965*	*0.96s*	*98.7 s*
Pipeline	*3738*	*1.14*	*1.88*	*3.11*	*4.33*	*58,209*	*3.27s*	*166.1 s*

6 Conclusions

A statistical methodology based on the DoE and RSM techniques with the aid of TCAD, has been utilized to study and analyze the effect of process variability in a 90nm bulk silicon technology; a similar approach can be used for SOI technologies, FINFET structures etc. and also for the more advanced technology nodes when process variation issues will be more problematic. Subsequently, a methodology for the architectural level modelling of the effects of process variation on propagation delay and leakage power was introduced. A statistical cell library has been built in order to provide both speed and efficiency in analyzing circuit delay performance. All the desired device parameters, whose variation specifications are obtained from the DoE and RSM analysis, are assumed to be Gaussian variables. A first-order canonical delay model is employed. The cell model can not only precisely reflect the process variation effect on delay and leakage power, but also can cope with different operating conditions and switching cases. This makes it easier to construct higher level blocks in the library with the help of SSTA and SPA. Based on the proposed methodology, the variation effects from manufacturing process can be analyzed at architectural level. A full analysis has been demonstrated on a 2-stage micropipeline circuit;

where it has been shown that this technique can achieve a comparable accuracy compared to a Monte Carlo simulation with the errors less than 5%, and save significant amount of computation time, at least 50 times faster than traditional SSTA and SPA approaches.

It is considered that the technique outlined in this paper, permits designers to efficiently assess the effects of variations in processing parameters, such as threshold voltage, together with supply voltage and operating temperature, on a design in terms of their potential impact on specification parameters such as propagation delay and leakage power early in the design cycle. Subsequently, the designer can choose which technology or cell library should be used to implement the design to ensure its robustness to the effects of process variation.

References

1. Kang, S., Leblebici, Y.: CMOS Digital Integrated Circuits, 2nd edn. Mc-Graw Hill (1999)
2. Montgomery, Design and Analysis of Experiments, 5th edn. John Wiley and Sons (2001)
3. Okada, K., Yamaoka, K., Onodera, H.: A Statistical Gate-Delay Model Considering Intra-Gate Variability. In: ICCAD, pp. 908–913 (2003)
4. Visweswariah, C., Ravindran, K., Kalafala, K., Walker, S.G., Narayan, S.: First-order incremental clock-based statistical timing analysis. In: DAC 2004, pp. 2170–2180 (2004)
5. Zhan, Y., Strojwas, A., Li, X., Pileggi, L., Newmark, D., Sharma, M.: Correlation aware statistical timing analysis with non-Gaussian delay distribution. In: DAC 2005 (2005)
6. Bhardwaj, S., Ghanta, P., Vrudhula, S.: A framework for statistical timing analysis using non-linear delay and slew models. In: ICCAD 2006, pp. 225–230 (2006)
7. Singhee, A., Singhal, S., Rutenbar, R.A.: Practical, Fast Monte Carlo Statistical Static Timing Analysis: Why and How. In: ICCAD 2008, pp. 190–195 (2008)
8. Rao, R., Srivastava, A., Blaauw, D., Sylvester, D.: Statistical Estimation of Leakage Current Considering Inter- and Intra-die Process Variation. In: International Symposium on Low Power Eletronics and Design, pp. 84–89. ACM, New York (2003)
9. Srivastava, A., Kaviraj, C., Shah, S., Sylvester, D., Blaauw, D.: A Novel Approach to Perform Gate-Level Yield Analysis and Optimization Considering Correlated Variations in Power and Performance. IEEE Trans. Computer-Aided Design 27(2), 272–285 (2008)
10. Chang, H.: Full-chip analysis of leakage power under process variations, including spatial correations. In: Proc. ACM/IEEE DAC 2005, pp. 523–528 (2005)
11. Ivan, E.: Sutherland "Micropipelines". Communications of ACM 32(6), 720–738 (1989)
12. Dierickx, B., Miranda, M., et al.: Propagating Variability from Technology to Architectural level. In: Workshop on IWPSD 2007, pp. 74–79 (2007)