Microelectronics Journal 69 (2017) 1–19

Contents lists available at ScienceDirect

Microelectronics Journal

journal homepage: www.elsevier.com/locate/mejo

Variability aware transistor stack based regression surrogate models for accurate and efficient statistical leakage estimation

Lokesh Garg

Department of ECE, Manipal University, Jaipur Dehmi Kalan,Kalan, Near GVK Toll Plaza, Jaipur-Ajmer Expressway, Jaipur, Rajasthan 303007, India

ARTICLE INFO

Keywords:
Standard cells
Leakage power
Process variation
Support Vector Machine
Probability density function

ABSTRACT

In this paper, we present highly accurate and very efficient stack based surrogate models for standby (idle-time) statistical leakage estimation of CMOS circuits using sampling based methods. Our aim is to replace SPICE simulation with our proposed surrogate models, which can be used to evaluate samples generated by sampling methods in variation space. Our methodology initiates by first characterizing the leakage of basic transistor stacks and then provides estimate of the leakage of basic gates based on these stacks. Transistor stacks are extracted across various standard cells, which are used to estimate leakage of CMOS gates with different input vector combinations. We develop Support Vector Machine (SVM) regression surrogate models and characterize the transistor stacks of CMOS gates, while accounting the combined effect of process variations in transistor length (L), threshold voltage (V_{th}), oxide thickness (T_{ox}), supply voltage ($0.6V - 1.2V$), temperature ($0°C - 100$°C) and width ($28nm - 200nm$), all scalable at the same time. For gates containing parallel transistor stacks, we merge the parallel transistors having identical inputs, which in turn allows us to use precharacterized basic stacks for leakage calculation, thus avoiding generation of new models. Our experiments illustrate that we only require 30 and 26 stack models to predict the subthreshold and gate tunneling leakage of 20 different gates across 176 input combinations, instead of characterizing leakage model for each input vector separately. SVM regression models generated in our approach have the ability to predict the leakage with maximum average error of 2.7% in mean (μ) and maximum average error of 3.1% in standard deviation (σ), both for OAI22 gate. Our results establish that there is on an average **10**× improvement in runtime while estimating the μ and σ of leakage of a gate within 10000 Monte Carlo simulation loop. Our approaches also result into a maximum of **221**× runtime improvement for C6288 ISCAS'85 benchmark circuit. We further develop Sparse SVM models using Support Vector spectrum pruning method, which reduces the runtime of the regression models with negligible increase in the error. Runtime efficiency of **17**× and **323**× is achieved on standard cell library and C6288 benchmark circuit respectively using Sparse SVM models. Our models outperform previous models based on analytical equations and Artificial Neural Network in terms of accuracy.

1. Introduction

Static leakage power is the main component in total power dissipation of digital circuits. Subthreshold leakage, gate tunneling leakage are major sources of static power dissipation. Due to lithographic limitations, parameters such as length, threshold voltage, oxide thickness tend to vary away from their mean value. In recent years, Dynamic voltage scaling (DVS) has been employed for performance optimization and power reduction of VLSI circuits. Temperature also affects the leakage power of a chip and can affect the different part of a chip in a different manner. Width must also be added to the model for prediction of leakage power of a gate for any transistor width to optimize the circuits. So, there is clearly a need to develop models that can predict the leakage power of a gate in a complete Process-Voltage-Temperature-Width (PVTW) space.

Gu et. al. [1] and Chen et. al. [2] have used the BSIM equations to model subthreshold leakage power. In [1], these equations are solved to calculate intermediate node voltages and leakage power while in [2], closed form of leakage is developed. Main drawback of both methodologies is that the effect of 'ON' transistors in 'OFF' network was not considered. Guiddi et. al. [3] calculate the leakage of a single minimum sized transistor a priori and then leakage of any gate is estimated by scaling this leakage power. Rao et. al. [4] assume the drain-source node voltages either at logic '0' value or '1'. Leakage of a gate is estimated by identifying the state of the transistor and then, adding the leakage of all transistors. Since the model used by authors is very conservative, hence, the error in leakage estimation is very high. Mukhopadhyay et. al. [5] model each transistor as sum of current sources (SCS). The equations

E-mail address: lokesh_garg20@yahoo.co.in.

http://dx.doi.org/10.1016/j.mejo.2017.05.015
Received 8 September 2016; Received in revised form 5 February 2017; Accepted 25 May 2017
0026-2692/ © 2017 Published by Elsevier Ltd.

used are very complex to calculate gate leakage and estimation methodology is slow too. Lee et. al. [6] characterize the subthreshold leakage of a single unit size transistor and then scale it according to the actual size of the transistor and stack size i.e. number of OFF transistors in a stack. This method is constrained and only allows equal size transistors on a stack. Yang et. al. [7] simplify the model presented in [6] by replacing the stack size and transistor size with a constant factor for each input vector of a gate. However, this method requires iterative method to calculate constant factors for varying stack factor and transistor size. Models presented in papers [1–7] were based on physical models which uses the device equations i.e. BSIM equations to estimate leakage of a gate. In presence of process variations, the calculation of mean of leakage requires integration of device equations which is very complex. The methodology was based on characterizing the leakage power through single transistor whether the equation is used for estimating the internal node voltages or used for scaling to calculate the leakage of gates. To remove this complexity, empirical models are used, in which BSIM equations are represented in the form of exponential linear (EL) or exponential quadratic (EQ) polynomial. Authors in [8] models the subthreshold leakage as an EL form and [9] as EQ model for every input vector which increases the number of models to be characterized. These models are regression based models which are generated by fitting simulation data to specific equation. These empirical models are simple analytical solutions which give large error in calculating subthreshold leakage in presence of large process variations [10]. Authors in [11] evaluate the accuracy of previous leakage models in [8,9,12] using BSIM4 transistor model [13]. State-of-the art methods were characterized using BSIM3 model [14]. More advanced BSIM4 model accounts several short channel effects, which were not considered in BSIM3 models. These physical effects impose high non-linearity into the model, which results in non-exponential linear behavior of leakage current. The average error in μ and σ can go upto ~ 20% and ~ 40% respectively. Recently, BSIM6 model has also been developed for more advanced technologies [15], which will further increase the non-linearity into leakage model due to secondary effects, It motivates to use regression based models to handle non-linearity introduced due to the use of BSIM4 models. We need to develop a kind of dynamic model without pre-assuming any kind of exponential or polynomial form.

Many EDA companies such as Cadence, Synopsys use.lib format to characterize leakage power, in which look-up tables are stored considering process parameters data at specific supply voltage and temperature value. Different look-up tables are developed to calculate leakage at different voltage and temperature conditions, thus limiting the leakage estimation at any generalized condition. Leakage characterization time may be less while not considering process variations but will be very high considering process variations. If a standard cell library consists of M gates and i^{th} gate has k_i inputs, then a total of $\sum_{i=1}^{M} 2^{k_i}$ subthreshold as well as same number of gate tunneling leakage models are required. It is very important to reduce the number of leakage models to reduce characterization time without sacrificing the accuracy. Authors in [16] developed a leakage contributor modeling concept, in which a set of contributors are characterized to reference all standard cells of considered standard cell library. For exp: leakage of a single PMOS transistor (contributor) can be used to provide leakage of NAND2, NAND3 and NAND4 gate with '1111' input vector, while standard.lib models characterize leakage using different look-up tables. The advantage of contributor models is reduction in the model characterization time due to less simulation of transistors and reduction in the memory used to store data for leakage models. Leakage contributor modeling resulted in 95× less transistor simulation and 5× reduction in characterization time. However, the contributor models presented in [16] were not thoroughly analyzed to provide subthreshold and gate tunneling leakage considering the effects of input vector combinations, thus may result in large error in leakage estimation using contributor models.

We present a comprehensive set of contributor models (called as stack models in our work) based on transistor stacks for subthreshold and gate tunneling leakage. For statistical leakage characterization of gates, it is more efficient to characterize different kind of transistor stacks instead of characterizing every gate for each input vector. Due to high non-linearity in PVTW space, surrogate modeling techniques such as Artificial Neural Network (ANN), Support Vector Machine (SVM) are more suitable while modeling non-linear performance parameters. This type of technique requires extensive samples in order to train the models. However once the models are trained, provide more accurate results than conservative models based on BSIM and empirical equations. In this context, Viraraghavan et. al. [17] have used the ANN model the leakage through a stack. The main drawback of this method is in the computation of scaling factors- only one scaling factor per input vector per gate is used for all PVT space. This may not be valid due to non-linear dependence of leakage power on PVT parameters. Maximum error in μ and σ reported is 20% for a gate in $130nm$ technology. The error would increase for circuits implemented in latest technology nodes as shown in Table 6.(b), further discussed in Section 4. The another limitation of model in [17] is that it does not use the width (W) of transistors on a stack as a input parameter for modeling stacks. ANN suffers from the condition of being trapped in local minima due to the use of gradient descent algorithm to calculate weights in the trained model. These models also suffer from the over learning i.e. high error for unseen data [18]. We added the width as another dimension to the leakage model. It increases the number of input parameters in the model but our model is still accurate enough. Leakage of a gate is function of width of transistors on a stack (W), global process (GP) and local process (LP) parameters, supply voltage (V_{DD}), temperature (T), input vector. Thus, leakage (I_{leak}) of a gate i with input vector v applied, can be given as:

$$I_{leak} = f_i^v(W_i, GP, LP_i, T, V_{DD}) \quad (1)$$

The addition of width in the models helps in reducing the number of characterized models other than basic stack models. These models are need to developed for the gates having parallel transistors or series of parallel transistors. We use SVM based performance surrogate modeling approach [19]. We propose a approach for efficient and accurate leakage estimation using smaller transistor stack models and also avoids the calculation of scaling factors (Column 5 of Table 6.(b)) as done in [17]. Thus, design efforts are reduced because of removal of scaling factors, and subsequently model characterization time for the stacks is reduced. Our proposed methodology is a generalized methodology is independent of technology and technology node, which may be applied to leakage characterization of non-planar devices i.e. FINFET, SOI [20]. The novel contributions of the proposed methodology are:

- Highly accurate stack based Subthreshold and gate tunneling leakage models are proposed. Our methodology requires less number of models for leakage characterization of 20 gates of CMOS standard cell library across 176 input vector combinations in complete PVTW space.
- Proposed approach combines the parallel transistors with same input and replaces it with a single transistor of effective width. Then the precharacterized basic models can be used, thus requires less number of models for the leakage calculation as compared to [17].
- Effective width calculation methodology is developed for the parallel transistors in CMOS gates with and without considering process variations, resulting into higher accuracy than the previous models [21].
- SVM regression based approach is employed for modeling leakage for such a large space, which requires less characterization time and is more accurate than look-up table and analytical equation based techniques.

The rest of the paper is organized as follows. Section 2 presents our

Download English Version:

https://daneshyari.com/en/article/4971129

Download Persian Version:

https://daneshyari.com/article/4971129

Daneshyari.com