INTEGRATION, the VLSI journal 43 (2010) 353–364

Contents lists available at ScienceDirect

INTEGRATION, the VLSI journal

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

Full-chip leakage analysis for 65 nm CMOS technology and beyond

Jiying Xue *, Tao Li, Yangdong Deng, Zhiping Yu

Institute of Microelectronics, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 31 October 2009
Received in revised form
16 April 2010
Accepted 11 May 2010

Keywords:
Statistical analysis
Leakage modeling
Process variation
Full-chip leakage distribution

ABSTRACT

This work proposes a full-chip leakage analysis framework for 65 nm technology and beyond. Analytical models are first constructed to capture the impact of process parameters on leakage current. Then a methodology is introduced to characterize leakage-related process variations in a systematic manner. On such a basis, an efficient procedure is developed to analyze the state-dependent power dissipation due to leakage of a large circuit block by taking into account different leakage mechanisms. Unlike many traditional approaches that rely on log-normal approximations, the proposed algorithm applies a quadratic model of the logarithm for the full-chip leakage current. It is able to handle both Gaussian and non-Gaussian parameter distributions. The model is validated with test chips manufactured with a commercial 65 nm CMOS process. Validation results prove that the proposed modeling methodology could achieve a higher accuracy than that from existing methods. Moreover, a full-chip leakage analysis using the developed model can be orders of magnitude faster than a Monte Carlo based approach.

1. Introduction

Aggressive scaling of CMOS devices at each technology generation has been offering ever-increasing integration capacity and circuit performance. However, the off-state leakage has also been rising significantly with technology scaling [1,2]. Under 65 nm CMOS technology, leakage power increases to 50% of the total chip power and dominates the switching power [3]. And Leakage analysis serves as the foundation for many important design issues. For instance, leakage should be analyzed in a systematic manner so that sufficient information could be collected for a good power-gating solution. Another concern is that the leakage variation would lead to voltage variation on the power grid and thus influence circuit performance. Consequently, it is essential to accurately estimate leakage current before a chip is actually manufactured so that it can be optimized in the VLSI design process.

In current CMOS technology, the major leakage components include subthreshold leakage (I_{sub}), gate tunneling leakage (I_{gate}), and reverse biased junction tunneling leakage (I_{junc}) in drain–substrate and source–substrate junctions [4]. For a drawn gate length (L_{drawn}) before the 90 nm CMOS technology, the junction tunneling leakage current is typically very small, while I_{sub} and I_{gate} dominate leakage in a circuit [5]. As a result, there have been extensive studies on I_{sub} and I_{gate}. However, when L_{drawn} is below 65 nm, higher substrate doping density and "halo" profiles (implant of the high doping region near the source and drain junctions of the channel) are applied to reduce the depletion region width of the source–substrate and drain–substrate junctions. A lower depletion region width and high-doping density near the source–substrate and drain–substrate junctions can cause significantly large tunneling current through these junctions under high reversed bias [4]. On the other hand, beyond 65 nm technology, new techniques are increasingly applied to moderate the gate-oxide tunneling effect with high-k dielectrics. The better insulation of the gate from the channel now brings gate oxide leakage under control (Fig. 1). Hence, the accurate estimation of the total leakage current considering *subthreshold* and *junction tunneling leakage* is extremely important for designing CMOS circuits in the nano-meter regime.

For the success of the aggressive scaling, it is indispensable to implement MOSFET performance booster techniques such as high stress contact-etch-stop-layer and eSiGe [6]. However, these boosters also have an impact on leakage current, which is not covered in most of the previous publications. Moreover, in many works (e.g., [4,7]), leakage analysis is often performed on a single MOSFET. Nevertheless, actual layouts usually have stacked gate MOSFETs deployed in critical paths [6].

Moreover, the increase of statistical variations in process parameters has posed serious challenges to the nano-scale circuit design. The variations can cause larger magnitude and wider distribution of leakage [8]. Despite advances in resolution enhancement techniques, lithographic variation continues to be a challenge for sub-65 nm technology, by introducing a severe distortion of the rectangular gate shape [9]. At the same time, aggressive scaling has resulted in many non-lithographic sources

* Corresponding author.
E-mail address: xuejy06@mails.tsinghua.edu.cn (J. Xue).

0167-9260/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.vlsi.2010.05.002

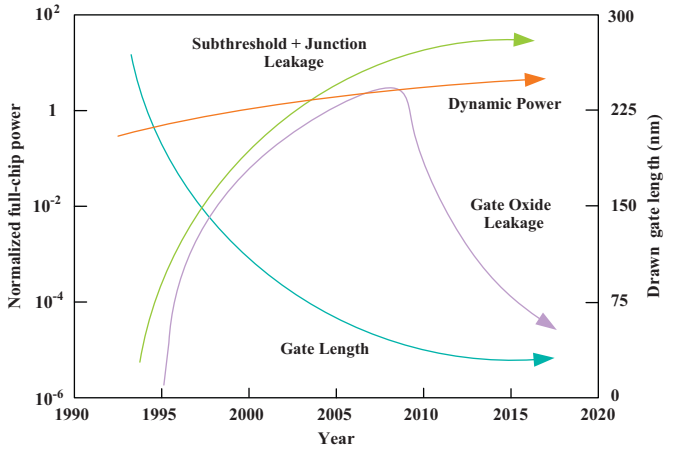

Fig. 1. Total chip dynamic and static power dissipation trends based on the International Technology Roadmap for Semiconductors [3]. Gate oxide leakage is under control by using high-k dielectrics [30,19].

of variations such as dopant variation, well-proximity effects, layout-dependent stress variation, and rapid thermal anneal (RTA) temperature induced variation [9]. In order to accurately model the process variations, both Gaussian and non-Gaussian parameter distributions should be characterized for improving model-to-hardware correlation.

Many works were conducted to perform a full-chip leakage analysis considering the process variations. However, most of them (e.g., [7,5,10–13]) either did not consider non-Gaussian parameter distributions, or did not include junction tunneling leakage that cannot be omitted for sub-65 nm CMOS technology. Recent work [14] took the above two issues into consideration. However, the leakage is approximated by a linear model, resulting in large errors for full-chip leakage analysis at sub-65 nm technology node, where the process parameters always vary by around 20% of their nominal values. Moreover, characterization of major variation sources for statistical leakage behavior has not been addressed yet. As a result, most of the previous methods cannot be directly used in leakage verification for manufacturing.

Based on our preliminary work [15], in this paper we propose a framework for full-chip leakage analysis for 65 nm technology and beyond. The framework consists of a complete flow that supports device-, gate-, and chip-level leakage analysis, as well as a methodology to characterize leakage-related process variations in a systematic manner. The concrete contributions of this work are as follows:

- Compact models for subthreshold and junction tunneling leakage are developed. The models consider short channel effect and narrow width effect. In addition, the models take layout parameters into account for the effects induced by stress techniques.
- A systematic characterization method for leakage-related parameter variations is proposed.
- A fast approach to analyze the state-dependent total leakage power of a large circuit block is established.
- Unlike many traditional approaches that rely on log-normal approximations, the proposed algorithm applies a quadratic model to the logarithm for the full-chip leakage current, and estimation accuracy of the proposed and conventional methods is compared.
- The proposed methodology is applied to a real SRAM design.

2. Statistical leakage modeling

2.1. Subthreshold leakage modeling

This section describes the general approach used to formulate the statistical model for subthreshold leakage current in a MOSFET. The model is developed from NMOS transistors, but can be easily extended to PMOS transistors.

2.1.1. Effective gate length model

In devices manufactured in 65 nm CMOS technology and beyond, the minimum feature size is much smaller than the optical wavelength, causing a severe distortion of the rectangular gate shape. As a consequence, there is a significant increase of the subthreshold leakage current (I_{sub}), while the current in the strong inversion region (I_{on}) is weakly affected [9]. This part we show how to translate an irregular gate structure into an equivalent transistor gate length, L_{eff}. Such a method is initially introduced by [9] and we add new parameters for better accuracy.

Fig. 2 shows a typical non-rectilinear gate profile. We can divide it into slices of different lengths and the same width W_0 along the width direction. Total leakage current for a given shape of gate can be approximated as the sum of currents over all the slices. Mathematically, the total leakage I_{tot} for a given gate of width W is expressed:

$$I_{tot} = \sum_{i=1}^{N} I_i(L_i, W_0) = I(L_{eff}, W) \quad (1)$$

Base on the exponential dependence of I_{sub} on L (further explained in Section 2.1.2), Table 1 summarizes the expressions of L_{eff}, which can be used in any circuit simulation tool for including the distortion of rectangular gate.

2.1.2. Modeling subthreshold leakage

In this part we develop a novel compact model for subthreshold leakage (I_{sub}). We have fabricated $\langle 1\,0\,0 \rangle$ NMOSFETs on the wafer of (1 0 0) substrate with drawn gate length varying from

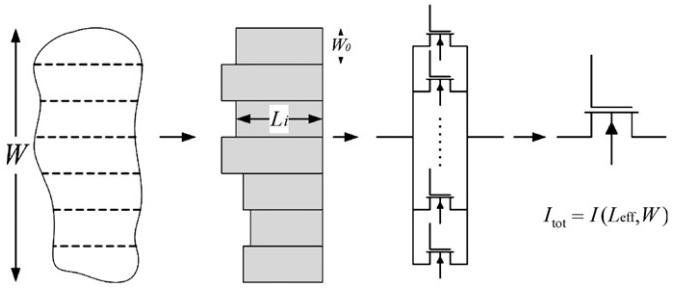

Fig. 2. Procedure to derive the effective gate length model.

Table 1
Summary of the effective gate length model.

Model parameters	Model expression
M	Total number of slices along the width direction
α	Fitting parameter
μ	$\mu = \frac{\sum_{i=1}^{M} L_i}{N}$
σ	$\sigma = \frac{\sqrt{\sum_{i=1}^{M}(L_i-\mu)^2}}{M}$
$L_{eff} = L_{min} + \alpha \cdot \ln\left(\frac{\sigma \cdot W}{W_0}\right)$	