Estimating MOSFET Leakage from Low-cost, Low-resolution Fast Parametric Test

Sharad Saxena, Takumi Uezono, Rakesh Vallishayee[2], Ron Lindley[1], Andrew Swimmer[1], Steve Winters[1]

PDF Solutions Inc., Richardson, TX 75082, USA
Phone: +1-972-889-3085, FAX: +1-972-889-2486, Email: sharad.saxena@pdf.com
[1]PDF Solutions Inc., San Diego, CA 92128, USA
[2]PDF Solutions Inc., San Jose, CA 95110, USA

ABSTRACT

A method of estimating the subthershold component of MOSFET off-state current (I_{offs}) using low-cost, low-resolution fast parallel parametric test is introduced. This method measures the subthreshold slope and uses it to estimate I_{offs}. Measurements of individual transistors show a very good agreement between measured I_{offs} and I_{offs} estimated using our approach. For a simple pad-efficient transistor array test-structure, where unselected devices can add additional noise to the subthreshold measurements, the sum of extracted I_{offs} for all transistors in an array is strongly correlated to the measured array I_{offs}, even though it does not match the measured array I_{offs}. The strong correlation is used to derive calibration factors which are then used to estimate individual transistor I_{offs} from array test structures. This allows statistical characterization of transistor leakage during volume production with minimal test time overhead. The applications of statistical off-state leakage characterization to diagnose IDDQ yield problems during production are also described.

INTRODUCTION

Leakage and variation are two undesirable consequences of transistor scaling. IDDQ is a critical parameter for most ICs, and ensuring high IDDQ yield requires characterizing not just the average values of the various components of leakage, but also the distribution of the leakage components.

The different sources of variation: systematic layout driven, lot-to-lot, wafer-to-wafer, die-to-die, across-chip and local variation all impact transistor leakage. Characterization and modeling the impact of these different sources of variation on leakage requires a large sample of measurements to separate the contribution of each source of variation on the different components of leakage.

Leakage monitoring during volume production requires fast measurements to minimize the test-time overhead. However, fast measurement of leakage poses special problems. Firstly, leakage currents are small, which implies that high-resolution measurements units are required. The expense of these high-accuracy units makes the use of parallel measurements to reduce test time cost-prohibitive. In contrast, moderate and strong inversion device characteristics like drive current (Idrive), threshold voltage (Vth) and transconductance (gm) can be accurately measured with low-resolution units allowing fast parallel test [1][2]. Secondly, accurate measurement requires long integration times.

This paper describes a method for fast estimation of transistor subthreshold current in off-state (I_{offs}) using low resolution measurement units. Gate induced drain leakage (GIDL) and gate current are the other main components of leakage in nanometer scale technologies. The technique of this paper does not allow the estimation of GIDL and gate current. They are estimated using separate test structures [2].

The I_{offs} estimation method makes use of I_d-V_g curves measured in sub-threshold. The sub-threshold currents are relatively larger than I_{offs}, allowing the use of low-resolution measurement units. Low-resolution measurements units in turn enable cost-effective parallel testing. When combined with our method for fast measurement of other devices characteristics (Idrive, Vth, gm, etc) [1][2], the I_{offs} estimation does not add any additional measurement overhead, it reuses the same I-V curves used to measure or extract the other parameters.

LEAKAGE CURRENT ESTIMATION METHOD

Drain-source current I_{ds} in subthreshold region is modeled by following expression assuming V_{ds} is larger than the thermal voltage [3]

$$I_{ds} = I_0 10^{(Vgs-Vt)/SS},$$

where I_0 is drain-source current when $V_{gs} = V_t$, V_{gs} is gate-source voltage, V_t is threshold voltage, and SS is subthreshold slope. From this equation, the plot of $\log(I_{ds})$ vs. V_{gs} is expected to be linear in sub-threshold region with a constant slope. By extrapolating $\log(I_{ds})$ vs. V_{gs} plot from the point where SS calculated, I_{offs} can be estimated as $\log(I_{ds})$-intercept. I_{offs} estimation method for nMOS can be summarized as follows:

978-1-4244-4259-1/09/$25.00 ©2009 IEEE

1. measure I_{ds} vs. V_{gs} curve for Vgs > 0;
2. calculate subthreshold slope (SS) by linear regression on consecutive sets of three log(I_{ds})-V_{gs} points;
3. extrapolate the linear regression to V_{gs}=0

Two modifications are made to the above procedure to account for the use of low-resolution measurements units and GIDL. In the first modification the SS is calculated from points where $I_{ds} > I_{res}$, where I_{res} is the resolution of the equipment. This has the additional benefit that when this technique is used to estimate leakage of transistors in a device array, the impact of the leakage of the non-selected devices in device arrays is minimized [1][4][5][6]. The second modification is to check for quality of fit (R^2) of the linear regression. If the R^2 is less than a threshold, the Id-Vg curve is rejected and I_{offs} is not extracted for such an abnormal device. An example of an abnormal device is a transistor that turns on before I_{res} is reached. Devices are also rejected if the slope of the regression is negative. Negative slope indicates devices with very large GIDL currents. These modifications allow a robust estimation of I_{offs} from low resolution measurement units.

EXAMPLES

The accuracy of the proposed method was assessed using measurements obtained from transistors fabricated in a 45 nm bulk CMOS process. First, the extrapolation method was applied to individual devices whose pins are directly connected to pads and are not shared with other devices. Figure 1 shows the comparison of I_{offs} measured directly on these structures and I_{offs} extracted from SS for a 45nm CMOS technology. As can be seen, a very good match is obtained between directly the measured I_{offs} and the I_{offs} estimated from SS. The correlation between extrapolated off current and measured off current is greater than 0.99 and the error of almost all of devices is less than 10%.

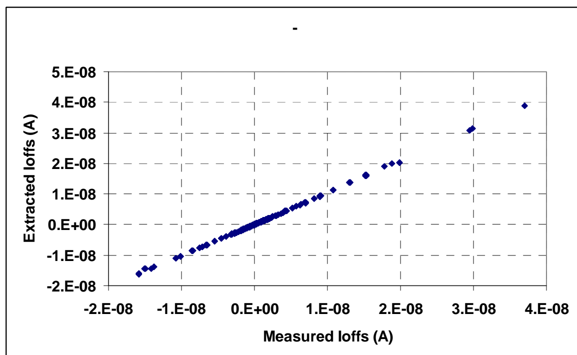

Figure 1: Comparison of I_{offs} estimated from SS extrapolation with direct measurement. The plot includes both nMOS and pMOS.

Transistor Arrays

Transistor arrays are compact and pad-efficient test structures for characterization of transistor variability [1][4][5][6]. This paper reports the results of applying I_{offs} estimation to a simple device array, which has been placed on scribe line of product wafers [1][2]. Figure 2 is a schematic of this test structure. In this transistor array 32 transistors share the source and drain pins. The source and drain pins are directly connected to the pads. Several arrays are implemented in one pad group to be tested in parallel. Figure 3 shows the distribution of I_{offs} extracted from parallel measurements made during volume production on device arrays placed on scribe line. The expected log-normal distribution of I_{offs} is observed.

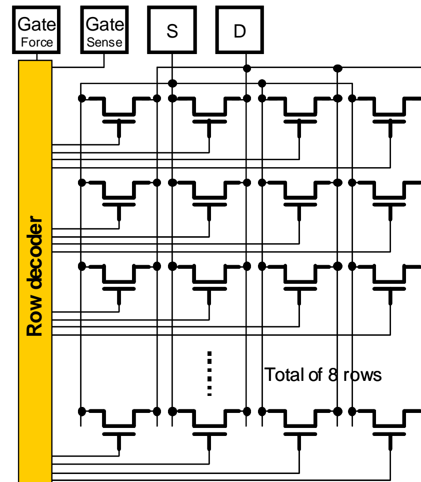

Figure 2: Schematic of transistor array test structure.

In addition to moderate and strong inversions characteristics of individual transistors in an array (eg Idrive, Vth, gm), this structure provides the measurement of I_{off} of the complete array on the source side. Array I_{offs} is the sum of the off-state current of all transistors in the array and is obtained by measuring the current on the source pad when no transistor is selected.

Figure 4 and Figure 5 shows the comparison of Array I_{offs} to the sum of extracted I_{offs} of the transistors in an array. Correlation is greater than 0.9 for pMOS and nMOS, wide and narrow transistors. However, the sum of extracted I_{offs} does not match the array I_{offs}. The likely cause of this discrepancy is the contribution of the non-selected devices in the array to the measured I_{ds}. Alternate device array architectures that minimize the contribution of non-selected devices to the measured I_{ds} could improve this match [4][5][6]. However, these techniques often require extra pads, which limit the number of arrays that can be put in a pad limited applications like scribe lines.

The high value of the correlation coefficient allows the extracted I_{offs} to be a good indicator of I_{offs}, for example to compare relative values across wafers and

across different splits. Moreover, the strong correlation makes it possible to estimate calibration factors which are used to calibrate the extracted I_{offs}. The next section describes the calibration procedure.

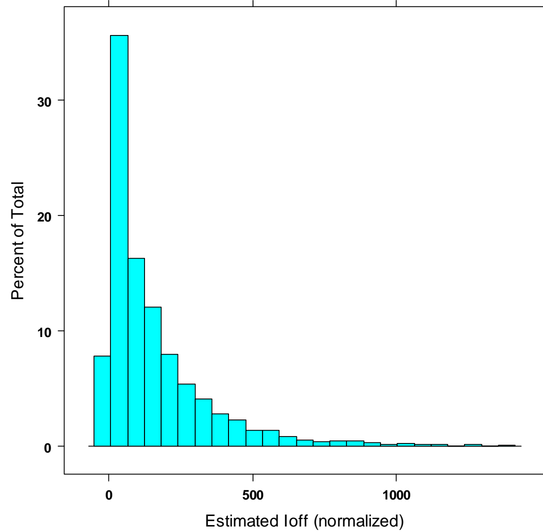

Figure 3: Distribution of extracted I_{offs} from device arrays measured on scribe line during volume production.

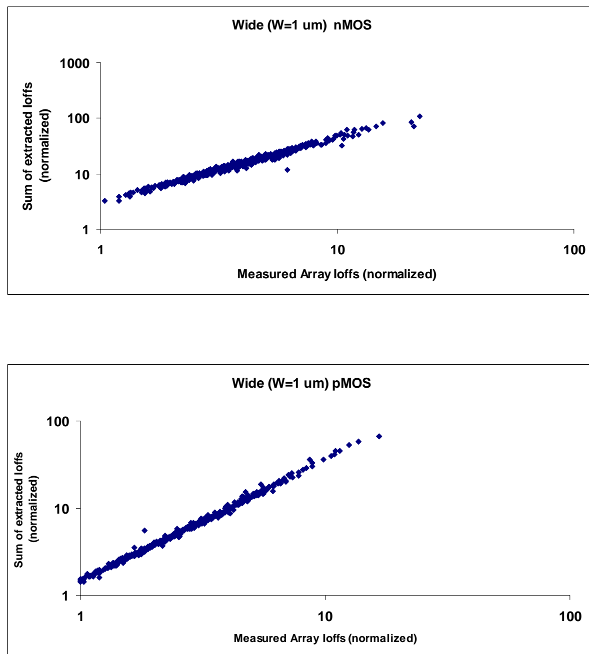

Figure 4: Correlation between sum of extracted I_{offs} of 32 transistor in an array with measured array I_{offs} for wide transistors. Both axes are normalized by the same same constant.

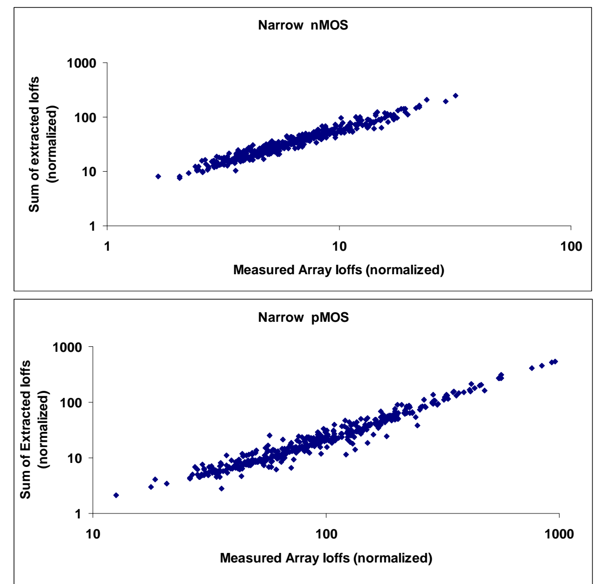

Figure 5: Correlation between sum of extracted I_{offs} of 32 transistor in an array with measured array I_{offs} for narrow transistors. Both axes are normalized by the same constant.

LEAKAGE CURRENT CALIBRATION

The sum of the extracted I_{offs} from the individual transistors in a transistor array has a very good linear regression with array I_{offs} for any given wafer (the multiple points for regression are the different sites on the wafer where the transistor arrays have been measured). However, this regression does not have slope 1.0 and intercept 0.0. It is possible to extract calibration factors from this slope and intercept for each individual transistor to correct for the contribution of unselected transistors.

Let us say that, for a specific array:

$$\sum_{i=1}^{32} Ioffs_i = Inter + \left(Ioffs_{Array} \times Slope\right),$$ where "i" refers to the individual transistors.

We correct each transistor I_{offs} using the formula:

$$Corrected_Ioffs_i = \frac{\left(Ioffs_i - \dfrac{Inter \times Ioffs_i}{\sum_{i=1}^{32} Ioffs_i}\right)}{Slope}$$

This assumes that the contribution of individual transistors to the intercept of the regression is proportional to the extracted I_{offs}. This calibration can be done for each wafer and each array, resulting in

robust I_{offs} estimates for each individual transistor. Figure 6 shows the comparison between measured array I_{offs} and the sum of extracted I_{offs} for each transistor. A very good match is obtained.

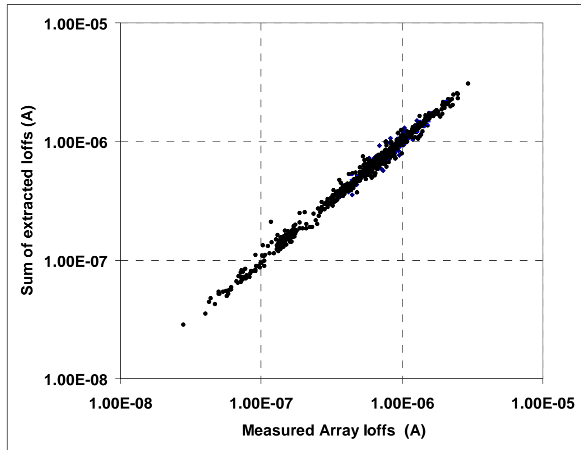

Figure 6: Comparison of sum of extracted I_{offs} after calibration with measured array I_{offs} for multiple device types and dimensions.

OTHER METHODS FOR LEAKAGE ESTIMATION

A number of test structures have been recently proposed for characterizing leakage variability [4][5][6]. These techniques focus on improving the accuracy of leakage measurements from transistor arrays, where the unselected transistor and selection circuits can contribute to the leakage measured on the shared drain or source pads. The focus of this paper is to make use of low-resolution measurement units for leakage estimation. The motivation for the use of low-resolution measurement units is to enable cost effective fast parallel testing. The leakage estimates obtained by the method described in this paper are suitable for monitoring, control and diagnosis during volume production rather than extremely accurate device characterization and modeling.

The method of extrapolating Id-Vg characteristics in subthreshold to extract I_{off} has also been reported before [7]. One application has been to extract subthershold slope and I_{offs} from Id-Vg curves in the presence of GIDL and parasitic STI transistors. This paper describes another application of the I_{offs} estimation by subthreshold slope extrapolation: cost effective fast parallel test.

APPLICATIONS

Fast parallel measurement during volume production allows a large sample of measurements to be taken, potentially every die and ever wafer, with minimal test time overhead. This enables rapid diagnosis and improvement of leakage related yield and performance loss.

Figure 7 and Figure 8 show one such application. I_{offs} extraction using the method described in this paper was performed during volume production on transistor array test structures placed in scribe line of a product manufactured in a 65nm technology. This large sample of measurements allowed spatial (across-wafer) analysis of I_{offs} distribution. Figure 7 shows I_{offs} distribution for various spatial zones for one of the nMOS transistors available in this technology. Zone 5 shows increase I_{offs} median and also increased variability in I_{offs}. Figure 8 shows product IDDQ for the same wafers. The IDDQ mean and variance is higher in zone 5 compared to other zones. The zonal correlation between I_{offs} and IDDQ helped isolate the causes of IDDQ variability and yield loss to transistor leakage rather than defects.

Figure 9 shows another application, the comparison of I_{offs} distribution of different transistor types offered in the technology. Different transistor types arise due layout and neighborhood differences, Vth, or Vdd differences. In this case, the examination of the distribution of extracted I_{offs} for different transistor types allowed identification of reasons for increase in the tail of the leakage distribution. Figure 9 shows that the distribution of I_{offs} for two transistors types available in this technology. An increase in the tail of the I_{offs} distribution is seen on wafer 5 for both types of transistors. This change in the distribution, which is made apparent by large sample of measurements on multiple transistor types, allows the isolation of the root-cause of leakage related yield loss in this product.

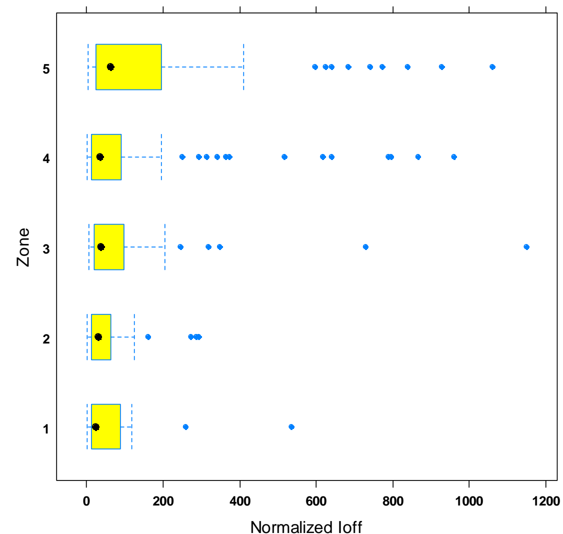

Figure 7: Across wafer spatial analysis of I_{offs} distribution. I_{offs} distribution for each spatial zone is shown as a box-plot.

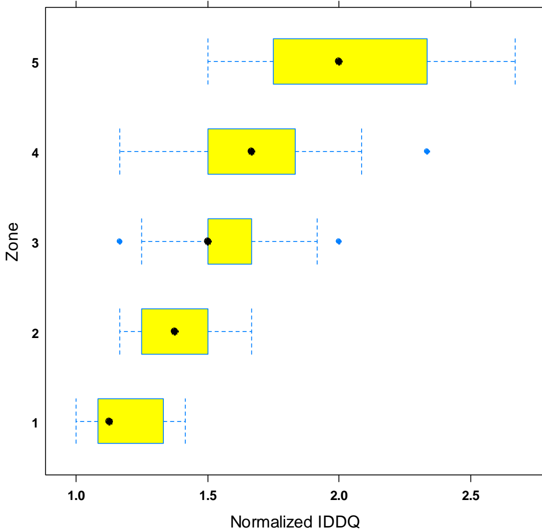

Figure 8: Across wafer spatial analysis of product IDDQ. IDDQ distribution of each spatial zone is shown as a box-plot.

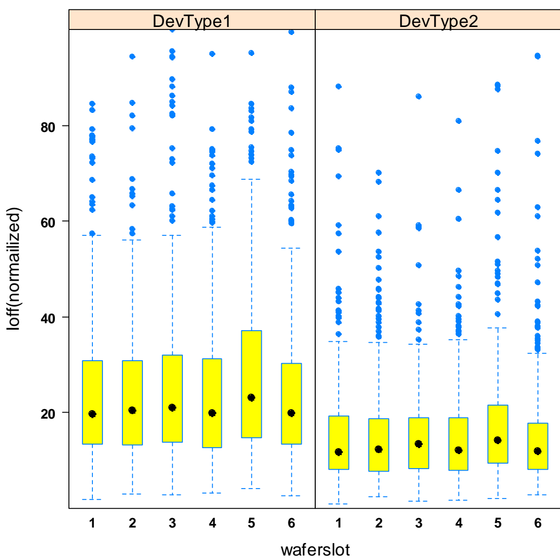

Figure 9: I_{offs} distribution of different device types available in a technology. I_{offs} distribution of each wafer is shown as a box-plot.

CONCLUSIONS

This paper described a method for using low-cost low-resolution parallel testers to extract off-state subthreshold leakage I_{offs}. I_{offs} was extracted from subthreshold slope. The method was applied to both individual and arrayed devices. In the case of individual devices, the extracted I_{offs} was in good agreement with the measured I_{offs} with error less than 10%. In the case of arrayed transistors, for a simple array architecture suitable for scribe line placement utilized in this work, the correlation between the sum of estimated off current and measured array off-current was greater than 0.9 even though estimated I_{offs} did not match array I_{offs}. The strong correlation was used to derive calibration factors, which were then used to estimate individual transistor I_{offs} from array test structures. The ability to estimate individual transistor I_{offs} using fast parallel test techniques allows statistical characterization of off-state currents enabling IDDQ estimation and yield improvement.

REFERENCES

[1] C. Hess, S. Saxena, H. Karbasi, et al. "Device Array Scribe Characterization Vehicle Test Chip for Ultra Fast Product Wafer Variability Monitoring", IEEE International Conference on Microelectronic Test Structures, pp. 145-149, March 2007

[2] S. Saxena, C. Hess, H. Karbasi, et al., "Variation in Transistor Performance and Leakage in Nanometer-scale technologies", *IEEE Transactions on Electron Devices*, vol. 55, num. 1, pp. 131-144, 2008.

[3] Y. Taur and T. H. Ning, "Fundamentals of Modern VLSI Devices," Cambridge University Press,1998

[4] T. Sato, H. Ueyama, N. Nakayama, et al.,, "A MOS transistor array with pico-ampere precision for accurate characterization of leakage current variation", IEEE Asian Solid-State Circuits Conferemce, pp. 389-392, 2008.

[5] B. Smith, U, Annamalai, A. Arriordaz, et al., "A novel biasing technique for addressable parametric arrays", IEEE International Conference on Microelectronic Test Structures, pp 166-171, 2008.

[6] K. Agarwal, F. Liu, C. McDowell, et al., "A test structure for characterizing device mismatches", IEEE Symp. on VLSI Circuits, 2006.

[7] H. Brut and R.M.D.A Velghe, "Contribution to the characterization of the hump effect in MOSFET submicronic technologies", IEEE International Conference on Microelectronic Test Structures, pp 188-193, 1999.