On-Chip Transistor Characterization Arrays with Digital Interfaces for Variability Characterization*

Simeon Realov, William McLaughlin, K. L. Shepard
Department of Electrical Engineering, Columbia University
E-mail: srealov@cisl.columbia.edu

Abstract

An on-chip test-and-measurement system with digital interfaces that can perform device-level characterization of large-dense arrays of transistors is demonstrated in 90- and 65-nm technologies. The collected variability data from the 90-nm run is used to create a statistical device model based on BSIM4.3 to capture random variability. Principal component analysis (PCA) is used to extract a reduced set of purely random variables from a set of correlated BSIM4.3 parameters. Different layout-dependent systematic effects, related to poly- and active-flares, STI-stress, and lithography limitations, are examined in both technologies. These layout-dependent effects are mapped to systematic shifts in BSIM4.3 and BSIM4.4 model parameters in 90- and 65-nm, respectively.

Keywords

CMOS, variability, modeling, statistical, PCA, measurement, on-chip, characterization, arrays, transistor

1. Introduction

Local, within-die process variability [1] is a critical concern in nanometer-scale CMOS as transistors are being affected by intrinsic fluctuations (dopant fluctuations, line-edge roughness) that cannot be controlled by the manufacturing process and as lithography and proximity effects, exacerbated by lithography techniques operating at resolution limits, have difficult-to-predict impacts on device characteristics. Traditionally, transistor compact modeling is done based on data collected from small sets of devices with pads which are individually probed. In the face of increasing device variability, this methodology becomes highly obsolete, as it makes it prohibitively expensive to examine the large number of devices needed to accurately model the increasing spread in device characteristics.

More recently, fairly comprehensive test structures based on ring oscillators have been developed to quantify delay variability[2,3]. The principal limitation of ring-oscillator-based approaches is that they "integrate" the effects of many device parameter variations (trioded and saturated device I-V characteristics, device capacitances), making it difficult to separate and characterize these in a way that would be amenable to statistical device modeling approaches and allow the results to be generalized to other circuit structures. Most recently, efforts have been focused on multiplexed transistor arrays [4] because they provide high-density access to a large number of devices for characterization. These designs have been limited by slow and difficult characterization through off-chip analog measurement and complexities associated with removing the effects of switch and parasitic resistances.

In this work, we describe the design of an on-chip characterization system that allows for the current-voltage characterization of large-dense multiplexed arrays of transistors. Digital interfaces allow these structures to be employed as part of an all-digital test methodology, and a Kelvin measurement approach eliminates the effects of on-chip parasitics. Implementation of the measurement circuitry using 2.5V thick-oxide devices allows for the seamless transfer of the on-chip characterization system from one technology node to the next, as is demonstrated by a successful migration of the measurement circuits designed in 90-nm to a 65-nm implementation. The details of the measurement arrays are described in Section II. Different choices for the array devices are made in the 90-nm and 65-nm arrays enabling different representative analysis of the data as described in Section III. Section IV concludes.

2. Test chip design

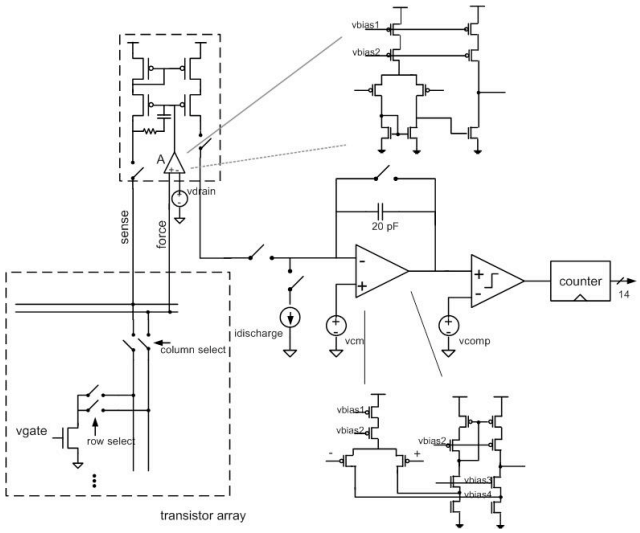

Figure 1: Schematic (simplified) of on-chip measurement circuits used for array characterization.

Building on our earlier design [5], Fig. 1 shows the block diagram of the on-chip measurement system. The die photo of the design as implemented in a 65-nm 1.0-V CMOS technology is shown in Fig. 2; this design is identical to the one implemented in a 90-nm 1.0V CMOS technology. In both cases, a custom layout generator is used to form the transistor array, containing 69 copies of each of 30 unique

* This work was supported by the SRC FCRP C2S2 Focus Center (www.c2s2.org) and by the SRC-GRC. We gratefully acknowledge UMC and Freescale Semiconductor for fabrication support.

transistors of differing sizes, layout styles and environments. The switches are implemented in thick-oxide transistors and consume < 40% of the total area of the array. All of the n-channel DUTs are placed in isolated p-wells biased by a dedicated power-supply pin in order to reduce the effects of substrate noise from digital switching activity.

Figure 2: Die photo of 1.3-mm-by-2.6-mm 65-nm CMOS chip. Metal fill obscures most of the noted features of the measurement circuits.

One device-under-test (DUT) is selected for measurement with row-select and column-select scan chains. The gate bias, *vgate*, is directly applied to the gate of the DUT. Separate force and sense lines typical of a four-point Kelvin measurement approach allow the current mirror to mirror the device current (*idrain*), while ensuring (through a negative feedback loop) that the drain voltage of the selected transistor is at *vdrain* regardless of any voltage drops across the switches and parasitic wire and via resistances. The cascoded current mirror is necessary to boost the output resistance of the mirror and avoid significant gain errors. The "forcing" amplifier is a simple two-stage design (Fig. 1). The gain of the entire feedback loop, which includes the DUT, varies from 55 to 80 dB depending on the operating point of the transistor under test.

All of the measurement circuits are implemented using thick-oxide devices operating at a 2.5-V supply, which ensures adequate headroom to characterize 1.0-V devices and allows easier migration to new technology nodes. The thick-oxide devices are not generally scaled and maintain relatively consistent design rules. In our own case, we were able to migrate the measurement circuits from a 90-nm process to a 65-nm process without any significant modifications to the design or layout.

A current-mode dual-slope integrating data converter is used to digitize the measured DC current levels entering the drain terminal of the DUT. The integrating amplifier has a single-stage folded cascode design used to provide an intrinsically stable high-gain, load-compensated frequency response. The high-gain comparator is implemented with an identical folded-cascode amplifier augmented by a second common-source stage. The on-chip digital counter, synthesized using the native high-speed digital devices, can operate up to 2 GHz. On-chip digital controllers programmably generate the required clocks and switch controls. An integrating capacitor value of 20 pF (implemented with a multilevel fringe capacitor) is used to accommodate the full current range of 0.5 mA to 50 nA. The ADC conversion has a linearity which exceeds 8 bits for all input current ranges and has 7 bits of absolute accuracy including the gain errors due to the mismatch and finite output resistance of the current mirror. At the typical current levels measured, the sample time needed for a single-point measurement is less than 100µs, which allows for a measurement throughput that easily exceeds 1000 data points per second, including the overhead necessary for synchronization between the test chip and the external measurement data acquisition hardware.

The characterization system allows for calibration of gain and offset errors. An empty cell in added to each row of the DUT array, enabling calibration of comparator offset and background leakage currents through the "off" switch transistors. Gain errors can be calibrated through the current-voltage characterization of a known off-chip resistor. Gain errors, however, are generally less than 5% because of the good matching and high output impedance of the current mirrors.

3. Measurement results and analysis

The measurement results and analysis presented in this section include data gathered from both the 90- and 65-nm test arrays. This represents the kind of analysis possible with the large quantities of device data produced by such measurement arrays. In all cases, we represent the I-V characteristics with BSIM4 model parameters, extracted from the measurement data. BSIM4.3 is used to model the 90-nm devices, while BSIM4.4 is used to model the 65-nm devices.

In the 90-nm arrays, devices of different sizes but identical layout environments enable the development of BSIM4-based statistical compact models to analyze the variability data from the 90-nm run, modeling the geometry dependence of random variability. Test structures to examine layout-dependent systematic effects enable poly- and active flares to be examined in the 90-nm array and layout dependent effects related to active-area dimensions, including STI stress, to be examined in the 65-nm array. In all cases, we seek to consider local, in-die process variability; therefore, all statistics reflect the transistor copies on a single die.

Table 1: BSIM4 parameter descriptions

BSIM4 Model Parameter	Description
VTH0	Long-channel threshold voltage at V_{BS}=0
U0	Low-field mobility
TOXE	Electrical gate equivalent oxide thickness
XL	Channel length offset due to mask/etch effect
ETA0	DIBL coefficient is subthreshold region
K1	First-order body bias coefficient

3.1. BSIM4 model parameter extraction

BSIM4 model fits are performed allowing only six parameters (given in Table I) to vary in the fit. VTH0, U0, TOXE, and XL capture variability in strong inversion (both triode and saturation) while K1 and ETA0 strongly model

device variability in the subthreshold region of operation. The starting point for the parameter fits in both cases is the BSIM4 model provided by the foundry, allowing relatively fast convergence in a few iterations. Aurora is used to perform the regression fits; typical rms errors are generally in the range of 2-4% with slightly larger fitting error (4-5%) observed in the 65-nm devices.

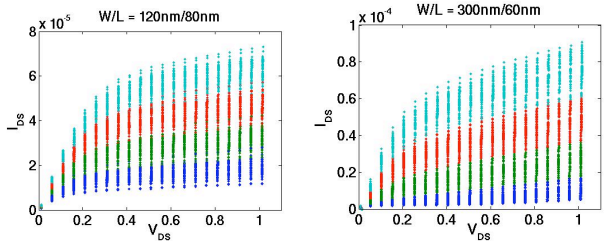

Figure 3: I_{DS} vs. V_{DS} strong inversion measurement results with V_{GS} = 0.5, 0.6, 0.7, and 0.8V for 69 identical devices. (Left) 120nm/80nm device in 90-nm run (Right) 300nm/60nm low-power (LP) device in 65-nm run

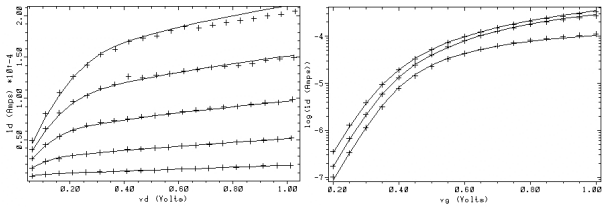

Figure 4: Example model fits for a W/L=500nm/80nm device. Measured data is shown with the crosses. (Left) Strong Inversion: V_{GS}=0.4, 0.5, 0.6, 0.7, and 0.8 V; rms error = 2.02%. (Right) Subthreshold: V_{DS}=0.1, 0.5, 1.0 V; rms error = 0.48%.

Fig. 3 shows examples of strong inversion measurements for 69 nominally identical devices in both the 90- and 65-nm CMOS technologies. The amount of variability in both cases is considerable. Fig. 4 shows representative measured I-V characteristics for a W/L = 500nm/80nm device in the 90-nm CMOS technology along with the associated regression fit to a BSIM4.3 model.

3.2. 90-nm array measurement results and analysis

The 90-nm arrays contain devices of different sizes in identical layout environments enabling statistical analysis of random variability. We use principal component analysis (PCA) to extract a reduced set of four principal components [6] from the set of six correlated model parameters described above. The four extracted principal components are uncorrelated by definition, and capture more than 95% of the original parameter correlations.

As a first step, the BSIM4 parameters, P, are normalized to P' according to

$$P' = \frac{P - \mu_P}{\sigma_P}, \qquad (1)$$

where μ_P and σ_P are the mean and standard deviation of the parameter P, respectively. Normalization of the model parameters is essential as it facilitates the extraction of the dominant eigenvectors based on their absolute contribution to the overall parameter correlations, irrespective of the actual mean values and standard deviations of the parameters in question.

The covariance matrix of the normalized set of parameters $\boldsymbol{C}$ is computed, along with its eigenvectors and eigenvalues. The eigenvectors of the covariance matrix are the principal components for the normalized parameter space, and their corresponding eigenvalues indicate the relative contribution of each principal component to the overall parameter correlation.

With the choice of four principal components, the statistics of the normalized parameters, $\boldsymbol{P'}$, can be modeled by a set of four independent normal random variables, $\boldsymbol{R}$:

$$(\boldsymbol{P'})^{\boldsymbol{T}} = \boldsymbol{E}\times\boldsymbol{E}^{\boldsymbol{T}}\times\boldsymbol{R}^{\boldsymbol{T}} \qquad (2)$$

where $\boldsymbol{E}$ is a matrix, the columns of which are the four principal components of the normalized parameter space (dominant eignenvectors of $\boldsymbol{C}$).

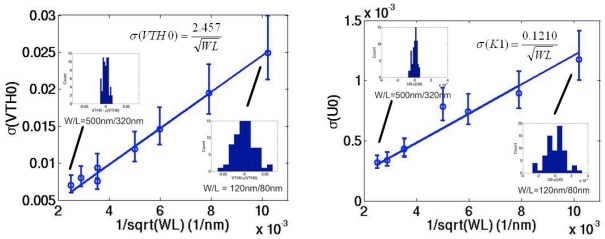

Figure 5: Geometry dependence of the standard deviation of model parameters VTH0 (Left) and U0 (Right). The standard deviations are plotted along with the 95% confidence intervals. Insets show the histograms of the distributions for the two endpoint geometries in each case.

In order to recover the original unnormalized parameter (P) from the randomly generated normalized parameter (P') using (1), the mean (μ_P) and standard deviation (σ_P) of P also need to be modeled. Under the assumption that the mean value of the parameter remains unchanged, the model for μ_P is a constant derived by averaging the population means across geometry. The model for σ_P , on the other hand, is an inverse function of the square-root of the device area according to the generally accepted model described in [7]. Fig. 5 shows the geometry dependence of the standard deviation of VTH0 and U0 as a linear function of $1/\sqrt{WL}$ for a set of eight devices with different lengths and widths. Simple linear regression fits are used to determine the values of the proportionality constants, A_{VTH0} and A_{U0}. This analysis applies to the entire set of six BSIM4 parameters.

Fig. 6 shows the correlated distributions of parameters K1 and TOXE as well as parameters U0 and VTH0 for a representative W/L=120nm/80nm transistor. These graphs include measurement data as well as the results of a Monte Carlo simulation of the PCA-based BSIM models developed here. The projections of the principal components onto the two-dimensional parameter subspaces are also shown. K1 and TOXE are strongly correlated (since K1 reflects subthreshold slope which is strongly dependent on C_{ox}) with a correlation that almost perfectly aligns with one of the principal components. U0 and VTH0 show weaker positive

correlation, which could be traced to their mutual dependence on substrate doping.

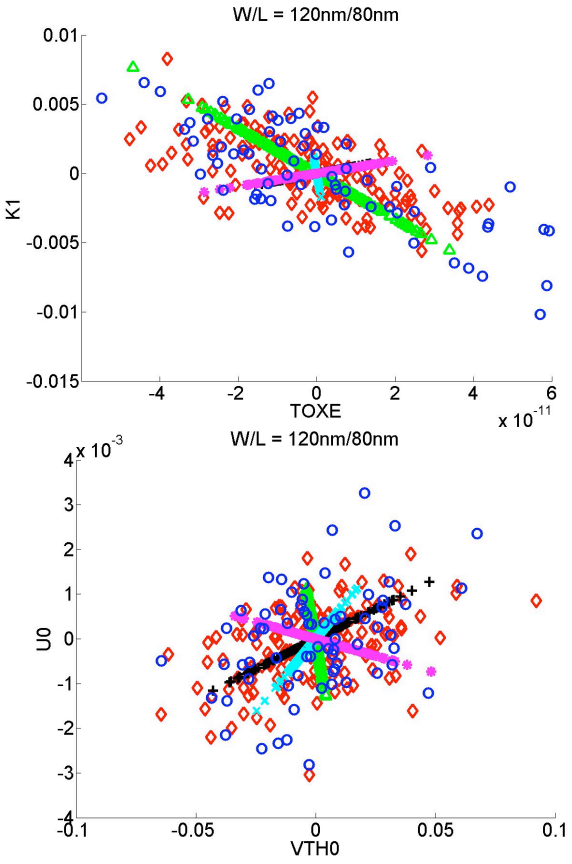

Figure 6: Correlated distributions of model parameters. The projections of the four principal component axes onto the associated plane are shown. The blue circles represent measured data while the red diamonds represent the results of Monte Carlo simulation of the statistical model developed here. (Top) K1 and TOXE. (Bottom) U0 and VTH0

In addition to developing a PCA-based statistical model for Monte-Carlo simulation, different layout-dependent systematic shifts in device behavior are also examined. Fig. 7 shows the mean and variance of two selected model parameters for active-flare and poly-flare layout topologies. XL is examined for poly-flare devices because lithography effects are expected to widen L for these geometries. A t-test

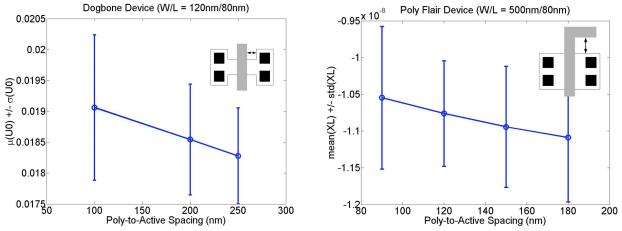

Figure 7: (Left) Layout dependence of U0 parameters for active flare layout style as a function of poly-to-active spacing. (Right) Layout-dependence of XL parameter on poly-to-active spacing for a poly flare layout style.

ascribes approximately 87-89% statistical significance to the differences between adjacent spacings and a 99% statistical significance to the two extreme spacings. For the active-flare topology, a width effect is expected, which should be reflected in model parameters that strongly determine I_{on} such as U0. More than 95% statistical significance applies in this case to adjacent spacings. The other layout-dependent systematic effects considered (which include dummy poly and contact spacing) are also weak, contrasting with results in other, possibly less mature, 90-nm technologies[3].

3.2. 65-nm array measurement results and Analysis

The 65-nm test array contains device structures (at the request of our foundry partner) that examine the effects of active-area topology and environment on device characteristics through both STI stress effects and subwavelength lithographic effects. Fig. 8 (left) shows the mean XL, along with the 95% confidence intervals, for a set of devices with W=1000nm and L=60nm, and varying spacing between the STI edge and the device gate. The difference between the mean values is statically significant at the 95% level, and indicates that the channel length offset decreases as the poly-to-STI-edge spacing decreases. In addition to this, Fig. 8 (right) shows the variation of VTH0 within the same set of devices. This parameter variation is also statistically significant at the 95% level, but only when comparing the two extreme values, and is probably related the effect of channel stress on the threshold voltage.

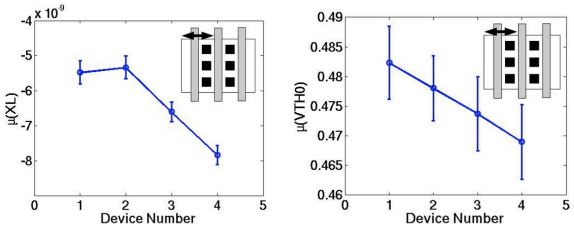

Figure 8: Mean values and 95% confidence intervals of XL (Left) and VTH0 (Right) for devices with W/L = 1000nm/60nm, and spacing between the gate and the STI edge of 180nm, 250nm, 480nm, and 2400nm for device number 1, 2, 3, and 4, respectively.

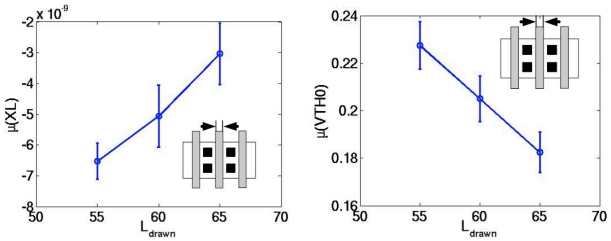

Figure 9: Mean values and 95% confidence intervals of XL (Left) and VTH0 (Right) for devices with W=300nm, and L = 55nm, 60nm, and 65nm.

Fig. 9 (left) shows the mean XL along with the 95% confidence intervals for three devices with equal widths (W=300nm), but different lengths (L=65nm, 60nm, and 55nm). A t-test performed between the three sets of measurements indicates that the observed results are

statistically significant at the 95% confidence level. The data suggests that as the drawn length is decreased close to the lithographical limitations of the technology, the effective channel length offset increases, and the proportion of drawn to effective channel length grows. Fig. 9 (right) examines the systematic shift of VTH0 for the same set of devices. Again, a t-test ascribes statistical significance to the observed results at the 95% confidence level, indicating that the effective threshold voltage increases as the channel length decreases, probably due to the reverse-V_T-roll-off effect observed in technologies which employ halo-doping.

4. Conclusion

A digitally-interfaced, on-chip transistor variability characterization system has been demonstrated in both 90- and 65-nm advanced CMOS processes. This system can be used to provide rapidly device-level current-voltage characteristics of a large number of devices in dense arrays. As an example of the kind of analysis possible with the volume of data acquired from these arrays, we have demonstrated the development of PCA-based statistical variability models for random, within-die variability and analysis of various layout-dependent systematically effects. Future work will include making these characterization arrays compatible with in-line testing, perhaps with scribe-line compatible layouts, as well as expanding the characterization capability to include C-V measurements of nominally-sized devices.

5. References

[1] S. Nassif, ISSCC 2000, pp. 368-369.
[2] N. Drego, et al, ISQED 2007, pp. 281-286.
[3] L.-T. Pang and B. Nikolic, VLSI Circuits 2006, pp. 69-70.
[4] K. Agarwal, et al, VLSI Circuits 2006, pp. 67-68
[5] V. Wang and K. L. Shepard, Elect. Lett. **43**, 2007. pp. 806-807.
[6] C. Michael and M. Ismail, JSSC **27**, 1992, pp. 154-166.
[7] M. Pelgrom, et al, JSSC **24**, 1989, pp. 1433-1440