An Array-Based Chip Lifetime Predictor Macro for Gate Dielectric Failures in Core and IO FETs

Pulkit Jain, *John Keane and Chris H. Kim
University of Minnesota, Minneapolis, MN 55455, USA
*Portland Technology Development, Intel Corporation, Hillsboro, OR 97124, USA

Abstract- A comprehensive Chip Lifetime Predictor (CLIP) macro for automatically characterizing gate dielectric failure reduces the stress time and silicon area by a factor proportional to the number of FETs to be tested. A flexible DUT cell that can be stressed in isolation without thicker $t_{ox}$ FETs to 4 times supply voltage, enables accurate lifetime prediction under different ON and OFF state dielectric breakdown modes for both low voltage core and high voltage IO devices.

I. INTRODUCTION

Device reliability mechanisms such as bias temperature instability, hot carrier injection, and Time Dependent Dielectric Breakdown (TDDB) have become pressing concerns in scaled technologies. While parametric shifts due to the former two can be mitigated using frequency guardbanding or circuit adaptation [1-2], such techniques are ineffective against the more catastrophic TDDB where even a single instance in a chip can cause an outright system failure. Fig. 1 shows the different TDDB modes affecting common digital circuits. While the on-state TDDB is most severe and conventionally assumed to be critical due to the entire gate area being exposed to stress, High Drain, High Source (HDHS) and High Drain (HD) OFF-state modes [3-4] might lead to earlier failure in circuits such as SRAM access devices that are exposed to an off-state stress for most of their lifetime. As for Input-Output (IO) devices, Electrical OverStress (EOS) and ElectroStatic Discharge (ESD) are of particular concern. Reliability margin targets for them become an issue with extensive use of high-voltage IOs and high-power CMOS devices at interface circuits in system on chips.

![Fig. 1 Different occurrence of gate dielectric failure. While ‘ON’ and ‘OFF-HD’ cases are most prominent, ‘OFF-HDHS’ is also seen in certain cases such as SRAM access devices.](image)

Particularly with TDDB, optimizing the fabrication process and using proper operating conditions based on accurate lifetime predictions is the most practical and effective approach. The main challenge is in the collection of massive statistical data from accelerated tests, as TDDB is a function of a number of variables including voltage, temperature, area, dielectric thickness, and purity (Fig. 2). Given the need for up to thousands of samples to correctly define a single Mean-Time-To-Failure (MTTF) value, traditional device probing quickly becomes cumbersome (Fig. 3). A previous characterization array for TDDB [5] only considered ON-state stress in core transistors which is not enough to obtain an accurate picture of system lifetime. A combined lifetime prediction methodology is needed to take into account different modes in tandem with their predicted time to failures. In this paper, we propose an array-based Chip Lifetime Predictor (CLIP) macro for efficiently collecting failure statistics under various accelerated stress conditions including ON-state and OFF-state stress modes for both low voltage core and high voltage IO devices.

![Fig. 2 Chip lifetime projection for TDDB based on accelerated stress involves mass data collection (e.g. up to 1000’s of samples per MTTF data) to make voltage, percentile, temperature, and area projections to actual product usage conditions.](image)

In the next section, we delve into the CLIP macro design and overall test strategy. Section III and IV describe the stress cell designs with measured statistics. In section V, the data is put together in perspective using the CLIP framework and finally we give a conclusion in section VI.
**II. MACRO DESIGN AND TEST STRATEGY**

**A. CLIP macro design**

The basic framework of the proposed CLIP macro is an array based statistical collection setup that can stress the DUTs in parallel while taking fast serial measurements controlled by a convenient scan-based interface (Fig. 4). This feature reduces the test time and test silicon area by a factor proportional to the number of DUTs. The gate terminal of the selected DUT is connected to the shared BL for IG measurements. The pre-charged BL gets discharged and the progressive TDDB in the form of IG is converted to a count by an on-chip current-to-digital converter.

**B. Current to digital conversion**

Reliability engineers employ both ‘hard’ and ‘soft’ increase in dielectric conduction for characterizing TDDB. We therefore employ two variants for current to digital conversion in this work. Fig. 6(a) shows the Current to Count Converter (CCC) to facilitate soft breakdown evaluation for the core case similar to the one used in [5].

**C. Calibration**

We embed replica stress cells, called “calibration cells” directly in the CLIP array (Fig. 4). These calibration cells were identical to the stress cells, but they did not have DUTs. Instead, a metal interconnect path was routed from the DUT gate node out to a pad. During calibration, a known range of resistances, R_{EXT} were attached to that pad in order to mimic a modes with programmable control. Two different flavors of flexible stress cells are needed for IO and core cases as will be discussed in the next section. As shown in the abstraction in Fig. 5, the underlying principle is to connect each DUT terminal to a stress voltage using on-chip switches rather than a hardwired inflexible connection. Flexible stress conditions used for the DUT cells have been tabulated in Fig. 5(c).

To avoid unrealistic GIDL behavior that may corrupt the stress data, a Voltage-Splitting technique (VST) was also implemented [3].

**B. Current to digital conversion**

Reliability engineers employ both ‘hard’ and ‘soft’ increase in dielectric conduction for characterizing TDDB. We therefore employ two variants for current to digital conversion in this work. Fig. 6(a) shows the Current to Count Converter (CCC) to facilitate soft breakdown evaluation for the core case similar to the one used in [5].
range of $R_{DUT}$, and measurements were run in the calibration cell. This isolates the $R_{DUT}$ from other extraneous resistances in the measurement path. The obtained results are shown in Fig. 7. The calibration cell also served useful as a marker cell during array operation.

Fig. 7 Calibration curves using the two current to digital converters. (a) CCC case, and (b) CBC case

III. CORE DEVICE BREAKDOWN CELL

A. Stress cell design
Details of the core transistor stress cell are shown in Fig. 8. The terminal voltages of the DUT are separately controlled by the STR/MEASb and VS/D/G control signals to provide programmable control of different flexible stress modes. If ‘FRESH’=0, the stress is gated off to prevent excess current during the long stress experiments. A timing logic selects cells in a manner that prevents over-shoot transients on the DUT and seepage of stress voltages to peripheral circuitry.

Fig. 8 Proposed DUT cell for core device breakdown. All FETs except DUT are thick tox devices.

B. Measured TDDB statistics
Adopting the methodology in Fig. 2, we first plot the results from different modes on a typical Weibull scale for straight line fits in Fig. 9(a-b). The rate of fails determined by the slope of this line, is related to [%] scaling. Fig. 9(a), shows the effect of pull down strength (or the breakdown threshold) used in CBC scheme. The ‘harder’ breakdown curves display a bend early in their evolution, and a lower slope.

Fig. 9(b) shows, that the two OFF modes have similar slopes and hence are projected to undergo almost identical [%] scaling. The MTTF versus voltage plot in Fig. 9(c) for different OFF and ON modes is used for [V] scaling. MTTFs for the different OFF modes were more than four orders of magnitude higher than ON mode at 4.5V. This difference is mainly attributed to the differences in effective gate area under stress. In ON mode, the entire gate area under the channel undergoes stress, while in the OFF mode it is limited to the overlap region. The [V] and [%] slopes for different cases are summarized in Fig. 9(d).

Fig. 9 (a) Effect of pull down strength in CBC scheme with VPDN=0.35V (b) Comparison of different OFF-modes (c) Relative voltage scaling in different modes (d) Relative comparison of [V] and [%] in different cases.

IV. IO DEVICE BREAKDOWN CELL

A. Stress cell design
The higher stress voltage (3-4 times the IO supply) and lack of a thicker tox device complicate the design of the IO stress cell shown in Fig. 10. A blocking circuit with dynamic biasing was designed to distribute the high stress, post-breakdown. A stack of two blocking circuits (single stack shown in Fig. 10 for simplicity) was sufficient to stress the cell upto 4 times nominal supply. Two flexible stress configurations were employed to provide OFF-state mode control.

Fig. 10 Proposed DUT cell for IO breakdown (top). No thicker tox devices are available so a blocking circuit was used to protect non DUT devices.

B. Measured TDDB statistics
Measured TDDB for a range of stress voltages is shown in Fig. 11 (left). [%] scaling from 63% to 1ppm for the core case is projected to be 100X larger than the IO case. Fig. 11 (right) compares the relative MTTF with different VSTR. A steeper slope is observed for the core case which translates into a 20X MTTF difference due to [V] scaling. MTTF for different
temperatures are shown in Fig. 12(a). Both core and IO FETs show Arrhenius trend in the measured regime. Spatial map shown in Fig. 12(b) of the individual cell’s TTF shows no obvious correlation.

Fig. 11 Measured breakdown data at different stress voltages for IO case

Fig. 12(a) MTTF for different temperatures. Both core and IO FETs show Arrhenius trend in the measured regime. (b) Spatial map of individual cell’s time to breakdown.

V. LIFETIME ESTIMATION USING CLIP MACRO

Fig. 13 shows the applied CLIP methodology for different stress profiles and gate types in tandem. The total gate areas for core and IO transistors were assumed to be 0.1cm$^2$ and 0.01cm$^2$, respectively. Duty cycle between OFF-state and ON-state is assumed to be 50% and a 1ppm failure percentile criterion was used. We observe that IO devices meet the lifetime requirement by a sufficient margin while the core transistor barely meets it. The chip microphotographs and summary of the core and IO CLIP arrays are given in Fig. 14.

VI. CONCLUSION

Optimizing the fabrication process and using proper operating conditions based on accurate lifetime predictions are the most practical and effective ways of dealing TDBD. However, the main challenge with this approach is in the collection of massive statistical data from accelerated tests, as TDBD is a function of a number of variables including voltage, temperature, area, dielectric thickness, and purity. In this work, we propose a CLIP macro for gate dielectric breakdown to reduce the stress time and silicon area by a factor proportional to the number of FETs to be tested. The essential part is a flexible DUT cell that can be stressed in isolation without thicker $t_{ox}$ FETs to 4 times the VDD, enabling accurate lifetime prediction under different ON and OFF state TDBDB modes for both low voltage core and high voltage IO devices.

Fig. 13 Comparison of projected lifetimes for IO and core devices for ON and OFF (avg. of HD and HDHS) states. Voltage, area, percentile, and temperature extrapolations (solid) are performed from measured statistical data (open).

Fig. 14 Test chip microphotographs of core and IO CLIP macros with chip summary.

ACKNOWLEDGMENTS

The authors would like to thank SRC and the Texas Analog Center of Excellence (TxACE) for financial support, and Dr. Vijay Reddy at Texas Instruments for technical discussions.

REFERENCES