JEDEC STANDARD

Gate Dielectric Breakdown

JESD263

MARCH 2024

JEDEC SOLID STATE TECHNOLOGY ASSOCIATION

NOTICE

JEDEC standards and publications contain material that has been prepared, reviewed, and approved through the JEDEC Board of Directors level and subsequently reviewed and approved by the JEDEC legal counsel.

JEDEC standards and publications are designed to serve the public interest through eliminating misunderstandings between manufacturers and purchasers, facilitating interchangeability and improvement of products, and assisting the purchaser in selecting and obtaining with minimum delay the proper product for use by those other than JEDEC members, whether the standard is to be used either domestically or internationally.

JEDEC standards and publications are adopted without regard to whether or not their adoption may involve patents or articles, materials, or processes. By such action JEDEC does not assume any liability to any patent owner, nor does it assume any obligation whatever to parties adopting the JEDEC standards or publications.

The information included in JEDEC standards and publications represents a sound approach to product specification and application, principally from the solid state device manufacturer viewpoint. Within the JEDEC organization there are procedures whereby a JEDEC standard or publication may be further processed and ultimately become an ANSI standard.

No claims to be in conformance with this standard may be made unless all requirements stated in the standard are met.

Inquiries, comments, and suggestions relative to the content of this JEDEC standard or publication should be addressed to JEDEC at the address below, or refer to www.jedec.org under Standards and Documents for alternative contact information.

Published by
©JEDEC Solid State Technology Association 2024
3103 North 10th Street
Suite 240 South
Arlington, VA 22201-2107

PRICE: Contact JEDEC

Printed in the U.S.A.

DO NOT VIOLATE
THE
LAW!

JEDEC Solid State Technology Association
3103 North 10th Street
Suite 240 South
Arlington, VA 22201-2107
https://www.jedec.org/contact

This page intentionally left blank.

GATE DIELECTRIC BREAKDOWN

Contents

Annexes

Contents (cont'd)

Figures

Tables

Foreword

This document describes procedures developed for estimating the overall integrity of gate dielectrics and for driving constant improvement in the gate dielectric manufacturing process in the integrated circuit manufacturing industry.

This document also describes robust techniques to detect breakdown and analysis methods for thin and ultra-thin gate dielectric films that exhibit large tunneling currents and soft breakdown.

The purpose of this document is to describe test procedures for estimating the reliability of gate dielectrics. It does not specify acceptance or rejection criteria for any of the described procedures.

The material contained within this standard is formulated under the cognizance of the JEDEC 14.2 Committee and approved by the JEDEC Board of Directors.

In addition to this standard, the guideline JEP194, *Guideline for Gate Oxide Reliability Robustness Evaluation Procedures for Silicon Carbide Power MOSFETs,* is available. It was formulated under the cognizance of the JEDEC JC-70.2 Committee and describes useful methods for GOX reliability assessment of thicker gate dielectrics which could be applied to SiC-based Power MOSFETs and also to Si-based Power MOSFETs.

This page intentionally left blank.

GATE DIELECTRIC BREAKDOWN

(From JEDEC Board Ballot JCB-23-58, formulated under the cognizance of the JC-14.2 Subcommittee on Wafer-Level Reliability.)

1 Scope

This document defines a constant voltage stress (CVS) test procedure for estimating time-dependent dielectric breakdown or "wear-out" of gate dielectrics. The test is designed to obtain area, voltage and temperature acceleration parameters required to estimate dielectric lifetime at use conditions and may be applied at wafer-level or on packaged devices.

The document also describes tests designed for simplicity, speed and ease of use; generally performed at wafer level: Voltage-Ramp (V-Ramp), Tunable V-Ramp, Current-Ramp (J-Ramp) and the Constant Current (Bounded J-Ramp) tests.

The test procedures defined herein can be used for capacitor testing, but are more commonly used for testing of transistors. Transistors may be tested in inversion or accumulation, and voltage bias may be applied to the gate or drain of the transistor. Since the extrapolation parameters may be different for inversion and accumulation, the product-relevant case (e.g. transistor in inversion) should be covered.

The document describes methods to obtain voltage, temperature, area and statistical-dependence for the time to reach a given failure fraction for a constant voltage stress or a regular unipolar AC-stress. These data can be used by simulation tools which estimate the effect of irregular voltage or temperature excursions on the lifetime of a transistor or capacitor. Such simulation tools are beyond the scope of this standard.

The document includes recommended data analysis methods and guidelines for statistical sampling and analysis as well as various sources of measurement error that could affect test results.

This document includes informative annexes that discuss test structure design (A.1), breakdown detection for CVS using an increase in noise criterion (A.2), breakdown detection for V-Ramp using a change in slope of the I-V curve (A.3), a methodology to compare CVS and V-Ramp data (A.4), statistical models (A.5 and A.6), methodology for calculating overall reliability (A.7), methodology for analyzing data which does not form a single-sloped Weibull distribution (A.8), a discussion of the interdependence of voltage and temperature acceleration factors for gate dielectric breakdown time (A.9), a comparison of AC and DC stress (A.10), a method for handling a range of dielectric thicknesses which occur in a process (A.11), a discussion of voltage acceleration in the Fowler-Nordheim tunneling regime vs. the direct-tunneling regime (A.12), a discussion of SiO_2 bandgap ionization by carriers with energies exceeding 9 eV (A.13), a discussion of transistor punch-through (A.14), a discussion of example failure rate calculations (B.1 through B.6), and Fowler-Nordheim and Direct-tunneling phenomena (C.1).

2 Overview

2.1 Stress Procedures

This document describes a constant voltage stress (CVS) test used to obtain acceleration parameters required to predict Time-Dependent Dielectric Breakdown (TDDB) or time to reach a given failure fraction of gate dielectrics used in advanced microelectronics.

The document also describes V-Ramp, Tunable V-Ramp, J-Ramp and Bounded J-Ramp tests used for process control of dielectric integrity.

It has been observed that breakdown in ultra-thin dielectrics can become “soft” or noisy. At the breakdown event, abrupt changes in current may not occur and device breakdown is difficult to detect. This document describes special methods for soft breakdown detection.

The document describes in detail the sources of error that can be encountered by test equipment, test structure properties, series resistance, environment, parasitics, and structural damage.

2.2 Choosing the Appropriate Stress Procedures

CVS and V-Ramp are most commonly used for dielectrics thinner than 10 nm. J-Ramp and Bounded J-Ramp are included in this document, since they are sometimes used for process monitoring.

CVS requires longer test times, and is typically used during process qualification and for process monitoring. The V-Ramp test can be used for rapid evaluation of dielectric quality and is often used during bring-up of a process technology to weed out extrinsic breakdown mechanisms and as a frequent monitor of a qualified process.

Prior to the stress, each test procedure includes a screening procedure to remove damaged devices. After the stress, a test is performed to confirm that a valid failure has occurred during the stress.

In CVS, the stress consists of a constant voltage applied until a failure or stop condition is reached. In order to determine operating points required by the CVS test, a pre-characterization test is performed.

In V-Ramp, the stress consists of a linear ramp in voltage from the use condition until a failure or stop condition is reached.

For both CVS and V-Ramp, it is common to periodically drop down to the operating voltage to measure the current in order to detect breakdown more easily.

In J-Ramp, the stress consists of an exponential ramp in current until a failure or stop condition is reached. In Bounded J-Ramp, the exponential ramp is stopped at a specified current density level and maintained there until a failure or stop condition is reached.

3 Terms and Definitions

The following symbols and terms are used in this document. They are organized in three sections and are listed alphabetically in each section for the convenience of the reader.

3.1 General Terms Applicable to All Tests Covered by this Standard

A_{diel} (cm^2): Dielectric gate area.

β: Weibull shape factor.

Compliance: The test equipment maximum current or voltage-forcing capability. Often the user can specify a compliance limit for a particular test. In this case the test instrumentation will not exceed this compliance limit during the test procedure.

Controlling variable: The applied variable in a stress procedure. For example, current in J-Ramp or voltage in V-Ramp and CVS are controlling variables.

DUT: device under test.

***E* (V/cm):** *E* is the estimated dielectric electric field. The general formula for E is: $E = V / th_{diel}$, where *V* is the voltage and th_{diel} is the estimated dielectric thickness, as determined by a consistent, documented method. The method (or a reference to the documented standard) must be included in the data report. In order to avoid confusion about different methodology associated with measurement of th_{diel}, voltage is more commonly used as a parameter than dielectric electric field whether using the power-law model or the exponential model of voltage/field acceleration.

Extrinsic Failure: An extrinsic failure for measured lifetime or projected lifetime is defined with a Weibull shape factor less than one, i.e. $0 < \beta < 1$. This corresponds to the situation in which the Weibull failure rate (Hazard rate) decreases with time. The extrinsic breakdown mode is caused by process-induced dielectric defects and the number of extrinsic failures can be reduced by improving the process.

I_{comp} (A): Compliance current; often the user can specify a compliance limit for a particular test.

I_{meas} (A): The measured dielectric current.

I_{use} (A): The typical measured current through the dielectric at the nominal use voltage.

I_{stress} (A): The dielectric current measured at the stress voltage or the current applied during a constant-current stress.

Intrinsic Failure: An intrinsic failure for measured lifetime or projected lifetime is generally defined with Weibull shape factor larger than one, i.e., $\beta > 1$. This corresponds to the situation in which the Weibull failure rate (Hazard rate) increases with time. For $\beta = 1$, this refers to a constant failure rate. An intrinsic failure is considered to be material-related and unavoidable even with process improvement. There are intrinsic failure cases with $\beta < 1$ due to specific material properties and or breakdown definition. For example, in the case of first soft breakdown of a dielectric with $th_{diel} \sim 1$ nm cases occur with $\beta < 1$.

3.1 General Terms Applicable to All Tests Covered by this Standard (cont'd)

L_g: The gate-length of a transistor

Q_{bd} (Coulombs): The accumulated charge passing through the dielectric prior to breakdown is defined as:

$$Q_{bd} = \int_{t=0}^{t=t_{bd}} I(t)dt$$

where t is time.

The Q_{bd} is calculated as the integral from t = 0 to t = t_{bd} where t_{bd} is the last measurement time at the step just prior to breakdown in a CVS or V-Ramp or the voltage measurement prior to breakdown in Bounded J-Ramp.

q_{bd} (C/cm^2): The accumulated charge density, passing through the dielectric at the detection of breakdown. Calculated as:

$$q_{bd} = \frac{Q_{bd}}{A_{diel}}$$

Response Variable: Refers to the variable which occurs in response to a controlling variable. For example, in J-Ramp, current is the controlling variable, and voltage is the response variable. In V-Ramp and CVS, voltage is the controlling variable and current is the response variable.

$t_{bd\_63\%}$ (s): The modal value of the Weibull distribution. In some cases, this is referred to as η.

t_{bd} (s): The recorded time at dielectric failure.

th_{diel} (nm): Thickness of a gate dielectric. In a data report which includes an electric field, the estimated dielectric thickness, as determined by a consistent, documented method must be included.

T_{use} (K): Use temperature – commonly, this is the worst-case sustained temperature during operation in the field. Alternatively, a single temperature which represents a temperature vs. time use-profile assumed when calculating the time to reach a given failure fraction.

V_{use} (V): The power supply voltage or nominal use voltage of the technology.

Weibull distribution: A Weibull distribution describes statistical phenomena such as dielectric breakdown in which time-to-breakdown (t_{bd}) is the random variable of interest even if dielectric thickness is perfectly uniform. A Weibull distribution or type 3 extreme-value distribution possesses the extreme- value character, which is most suitable for the characterization of area-dependence of dielectric breakdown.

3.2 Terms Specific to Voltage Stress

Drain-bias stress: Stressing a device to breakdown by applying a voltage to the drain of a transistor and grounding the source, gate and well.

γ_V: Voltage-related constant used in the V-model which assumes that $t_{bd\_63\%} \propto \exp(-\gamma_V Vg)$.

γ_E: Electric-field-related constant used in the E-model which assumes that $t_{bd\_63\%} \propto \exp(-\gamma_E Eg)$.

Gate-bias stress: Stressing a device to breakdown by applying a voltage to the gate of a transistor and grounding the source, drain and well.

I_{bd} (A): Dielectric current measured just after dielectric breakdown in a V-Ramp stress (response variable in a V-Ramp stress).

$I_{previous}$ (A): The measured dielectric current in a previous step of a stress.

I_{fail} (A): Specified failure current for breakdown detection – not the compliance value.

I_{sense} (A): Current measured at $V_{sense.}$

***n*:** or Voltage Acceleration Parameter, is the slope of the $ln(t_{63})$ vs. $ln(V_{stress})$ plot.

r_{crit}: Critical ratio between measured currents at consecutive time intervals in a CVS which indicates dielectric breakdown. If the ratio between measured currents at consecutive time intervals in a CVS exceeds r_{crit} (usually 2 to 10) the dielectric is considered to have undergone dielectric breakdown.

t_{fail} (s): The total time from initial stress until a current limit is reached.

t_{1bd} (s): The time for the first percolation path to form across a dielectric which results in higher leakage across the dielectric.

t_{pbd} (s): The time after t_{1bd} for a single percolation path across a dielectric to widen, allowing higher current flow until a current limit is reached.

t_{int} (s): The time that stress is applied before the stress is interrupted and current is measured at V_{use} during the stress interruption technique for detecting breakdown.

t_{meas} (s): The time interval for measuring the stress current (I_{stress}) during the constant voltage test.

t_{record} (s): The time interval for recording current measurements (I_{meas}) during the constant voltage test.

t_{wait} (s): The wait time after stress is interrupted before current is measured at V_{use} during the stress interruption technique for detecting breakdown.

V_{bd} (V): The voltage reached in the V-Ramp as breakdown is detected.

3.2 Terms Specific to Voltage Stress (cont'd)

$V_{bd\_63\%}$ (V): The modal voltage in a V_{bd} distribution.

V-Ramp: This test applies an increasing voltage ramp to a test structure until dielectric breakdown occurs or a compliance condition is reached.

V_{step}: The increase in voltage between consecutive steps of a V-Ramp.

V_{sense} (V): The voltage at which current is measured to detect breakdown before and after (and in some cases, during) CVS or V-Ramp.

V_{start} (V): The first voltage applied during a V-Ramp.

V_{stress} (V): The dielectric stress voltage.

3.3 Terms Specific to Current Stress

I_{init} (A): The initial current applied in a J-Ramp stress (current is the controlling variable in a J-Ramp stress).

I_{max} (A): The maximum current allowed in a J-Ramp stress.

J (A/cm^2): The dielectric current density calculated by dividing the dielectric current by the dielectric area (A_{diel}).

J-Ramp: This test applies an increasing current ramp (usually in logarithmic increments) to a test structure until dielectric breakdown occurs, I_{max} is reached or if I_{max} is not set, a compliance condition is reached.

r_{step}: ratio of current applied in successive steps in a J-Ramp

V_{meas}: the measured voltage across a dielectric at a given applied current.

V_{prev}: the measured voltage across a dielectric in a previous step of the stress.

4 Voltage Stress for Dielectrics (CVS, V-Ramp and Tunable V-Ramp)

4.1 Test Configuration

Although it is possible to perform dielectric stress on capacitors or transistors, transistors are commonly used either in inversion or accumulation with voltage bias on the gate or the drain.

It is important to minimize any voltage drops in the connection of the power supply to the terminals of the DUT due to series resistance that may lead to errors in the actual applied voltage.

The DUT should not have contacts or probe pads over it. Compressive stress caused by probing pads over dielectrics can result in device damage and inaccurate defect density measurements.

It is important to confirm that the parameters calculated from gate dielectric breakdown testing represent the nominal integrated circuit manufacturing process. It is recommended to use at least three nonconsecutive wafer lots – include at least three wafers from each lot, and test at least 30 devices from each wafer randomly distributed throughout the wafer. Parameters such as gate leakage, off-current, threshold voltage and interconnect series resistance are often measured in order to confirm that the wafers used to determine gate dielectric breakdown parameters are all within the specifications defined for a given process in the manufacturing facility. When process control monitors have been defined, it is important to ensure that all wafers used fall within the required process control specifications. For more guidance on wafer lots and sample size, please see JEP001-2, section 4. In addition, the sample size at each stress point should be carefully chosen to provide acceptable confidence limits. Please see [3] for calculating confidence bounds for Weibull distributions.

In addition, fabrication facilities often perform further checks of the process uniformity with respect to gate dielectric breakdown by first establishing a correlation between CVS and V-Ramp (see A.4) and then performing V-Ramp measurements on an additional quantity of identical test structures.

The extrapolated product lifetime must take into consideration the difference in area between the test devices and the product. One of the properties of Weibull statistics is a shift of the distribution with device area. Area scaling of time-to-breakdown, t_{bd}, can be defined as follows [2]:

$$\frac{t_{bd1\ 63\%}}{t_{bd2\_63\%}} = \left[\frac{A_2}{A_1}\right]^{1/\beta} \tag{4.1}$$

$t_{bd1\_63\%}$ is time-to-breakdown for device of area A_1 and $t_{bd2\_63\%}$ is time-to-breakdown for device of area A_2. β is the Weibull shape factor. The above equation is only valid if any two t_{bd} distributions measured with two different areas will merge together according to the following equation:

$$ln\big(-ln(1-F_2)\big) = ln\big(-ln(1-F_1)\big) + ln(A_2/A_1) \tag{4.2}$$

This ensures that the derived β value from the equation of area scaling of time-to-breakdown (t_{bd}) truly represents the Weibull shape factor.

Cases where t_{bd} distributions follow non-Weibull characteristics are discussed in A.8 and in A.11.

4.1 Test Configuration (cont'd)

In selecting the optimal test area, the following tradeoffs must be considered. Larger areas provide more information about the low-percentile failure rate. In state-of-the-art technologies, large test-area can only be achieved by connecting many transistors within a layout and stressing a common gate of the array. Hard dielectric breakdown, particularly from drain-bias stress, can result in failure of the interconnect to the broken gate dielectric within the array, so that the expected high current at breakdown is not observed and the test is continued without noting the breakdown (see A.1.1.1). In such a case, the area must be reduced in order to accurately detect breakdown.

In order to confirm that a given large-area test structure is not suffering from non-uniform voltage stress due to series resistance, an inability to detect breakdown due to local interconnect failure, or any other issues, some smaller-area structures should also be tested and Poisson-area scaling (see equation 4.1 and equation 4.2) should be performed. If the results from all of the areas follow Poisson-area scaling, then the integrity of all of the test structures is confirmed. If the largest areas do not follow Poisson-area scaling, they will have to be dropped or redesigned in order to obtain accurate t_{bd} or V_{bd} results.

A.1.4 covers area-dependence issues which can occur in capacitor testers (generally used only for gate oxides thicker than 10 nm) due to perimeter contributions such as gate-edge and isolation-edge effects and discusses how to design test structures to cover these failure mechanisms.

The optimal test area also depends on how sensitive a given circuit is to an increase in leakage. For example, some analog circuits are sensitive to a small difference in leakage between two transistors. In this case, very small-area test structures must be used. In contrast, for many SRAM, a leakage of >1 μA and for many digital circuits, a leakage of >10 μA can be tolerated. Since many products have large areas covered by SRAM or digital circuits, there can be a strong impetus to understand the failure statistics for failures that exceed 1 μA to 10 μA in leakage. For this purpose, as an example, if a small-area (e.g., 10^{-3} μm^2) test structure has a background leakage level that ranges from 1 pA to 10 pA, the area can be increased up to 100 μm^2 to 1000 μm^2 while preserving the ability to detect a failure with a leakage level of 10 μA (failure current would be 10× higher than the background leakage level). Although extrinsic and intrinsic failures both exhibit a range of failure currents, extrinsic failures with high leakage levels are most egregious for integrated circuits. As a process is optimized, it is common practice to use large-area test structures (~1000 μm^2) to detect extrinsic gate dielectric breakdown, and continue to tweak the process until such failures become very rare.

A.5 explains how to censor extrinsic failures in order to accurately model intrinsic dielectric wear-out.

For gate-bias stress which undergoes soft breakdown, and when using an I_{fail} criterion, testing larger arrays of transistors is recommended. The data may be analyzed as described in A.8. Alternatively, a simpler conservative approach may also be used. After combining multiple areas based on equation 4.2, and verifying that a single scaled distribution is formed, if that distribution does not form a straight line on a Weibull scale and follows a curvature as shown in Figure 10, the data from the largest area can be fit with a single slope which can be used as β_{fail}.

4.1 Test Configuration (cont'd)

Once the test structures have been confirmed to be valid through successful area-scaling and almost all extrinsic failures have been eliminated from the statistical distribution, the voltage acceleration parameter can be determined. In order to determine the voltage acceleration parameter, *n,* for power-law voltage acceleration or γ for exponential models, required to extrapolate product lifetime from accelerated stress conditions, the CVS test is performed at several voltages. Weibull statistics are recommended to fit the failure distribution. Since the voltage acceleration depends on the temperature, it is recommended to perform the voltage-dependent measurements at the use temperature. In some cases, it is convenient to use the temperature at which high-temperature-operating-life (HTOL) testing is performed.

Voltage-dependent CVS measurements are performed at voltages significantly higher than V_{use} in order to obtain reasonable t_{bd} (10 - 10,000 s, unless package-level testing is used). In cases where direct-tunneling occurs at V_{use}, for dielectrics with a very small thickness range, the tunneling mechanism at Vstress can be Fowler-Nordheim (F-N) (see Annex C). It has been found that the power-law exponent is smaller in the F-N regime (see A.12). Using a voltage-acceleration factor obtained from testing in the F-N regime results in a shorter projection of time to reach a given failure fraction. Performing the voltage-dependent CVS measurements at the highest temperature allowed for the technology allows the lowest-possible voltages to be used (more likely to be in the direct-tunneling regime). Many researchers have observed that the voltage acceleration factor tends to be lower at higher temperature, but higher in the direct-tunneling regime. Voltage-dependent CVS measurements at the highest temperature allowed for the technology are valid, and the effects on VAF can be understood by studying A.9 and A.12. Although more complicated, it may be possible to obtain more accurate lifetime estimates using a temperature-dependent voltage acceleration with equation A.9.1.

For transistors operating at voltages above 5 V, additional concerns include voltage reduction due to negative charge trapping during CVS. Please refer to the JC70.2 SiC Gate Dielectric TDDB Guideline. The information about CVS contained therein also applies to thick gate dielectrics in Si-based transistors. Impact ionization (SiO_2 bandgap ionization by carriers with energies exceeding 9 eV) is not typically a concern for CVS of gate dielectric thicknesses below 10 nm. Please also see A.13 for a discussion of impact ionization and its effect on breakdown of thick gate dielectrics.

4.2 Pre-stress Screen for CVS and V-Ramp

The purpose of the pre-test is to identify initial device failures.

1) Set and record stress temperature. Temperature should be controlled within ± 2 °C.
2) Determine typical I_{use} values at V_{use}.
3) Force V_{use}.
4) Measure I_{meas}.
5) If I_{meas} > 10× average value of I_{use} determined from the pre-characterization test, record as initial failure.

4.3 Stress Component for CVS

Devices that pass the pre-CVS test are then subject to the constant voltage test. In order that the compliance circuit in the test equipment does not interfere with the determination of breakdown it is recommended that the compliance of the test system (I_{comp}) be at least 10× greater than I_{stress} [5].

1) Apply or ramp the voltage to the stress voltage (V_{stress}) ensuring that voltage overshoots do not exceed 1% of V_{stress}. The voltage ramp time should be less than 1% of the anticipated time-to-fail (t_{bd}). The minimum value is 100 ms.

2) After the voltage ramp, the voltage is held at V_{stress} and the current is periodically monitored and logged to a data file. The current may be monitored at V_{stress} and also at V_{sense}. Setting V_{sense} at V_{use} allows the most sensitive breakdown detection capability. The current can be monitored at short time intervals (t_{meas}); however, the current should be recorded in the data file at time intervals (t_{record}). The record time interval should be less than 1% of the anticipated time-to-fail (t_{bd}). For practical considerations t_{record} can be evenly-spaced log-time intervals. The recorded data can be used to validate the actual breakdown event.

4.3.1 Stress Configuration

The following table provides guidance for the stress bias conditions at a given stress temperature applied to each wafer per lot. A minimum of 30 DUTs per stress bias condition need to be used. Table 1 is simply a guideline. It is important to ensure that the stress voltage range does not lead to over-stress conditions triggering different degradation mechanisms. At short stress times, the resolution of the electronics may lead to unacceptable uncertainty in determining $t_{bd\_63\%}$. The recommended area ratio between the largest and smallest areas is at least 100 (1st Area/3rd Area >100).

Table 1 — Stress Configuration Guidance for CVS

	1st Voltage: Voltage to achieve $t_{bd\_63\%}$ of ~10 s for 2nd Area	2nd Voltage: Voltage to achieve $t_{bd\_63\%}$ of ~100 s for 2nd Area	3rd Voltage: Voltage to achieve $t_{bd\_63\%}$ of ~1000 s for 2nd Area
1st Area: Area to achieve $t_{bd\_63\%}$ of ~10 s at 2nd Voltage		x	
2nd Area: Area to achieve $t_{bd\_63\%}$ of ~100 s at 2nd Voltage	x	x, stress at 2 additional temperatures	x
3rd Area: Area to achieve $t_{bd\_63\%}$ of ~1000 s at 2nd Voltage		x	

Recent literature confirms that time to gate dielectric breakdown follows a power-law model vs. voltage and this is widely used in the industry (see also JEP122). Using the power-law model, the lifetime at V_{use} can be estimated using the following relationship: $t_{bd\_63\%} \propto V^{-n}$ fitting the experimental datapoints at three voltages.

For single-sloped Weibull distributions, if area-scaling is performed as in 4.8.1, the dielectric lifetime at A_{use} can be estimated using the following relationship: $t_{bd\_63\%} \propto A^{-1/\beta}$, fitting the experimental datapoints at three areas.

4.3.1 Stress Configuration (cont'd)

In the case of drain-bias stress for short gate-length transistors, the lifetime depends on off-current (current flowing from the source to the drain when the drain is biased and source, gate, and well are grounded). Off-current is affected by threshold voltage and effective gate-length (for a given physical gate-length, effective gate-length can be reduced by e.g. lateral diffusion, selective epitaxial features and angled implants). Drain-bias CVS measurements should be performed on structures with highest off-current (lowest threshold voltage and shortest effective gate-length for a given technology). Area-scaling for drain-bias stress consists of increasing the width of a transistor or using an array with many transistors of a given gate-length.

Under accelerated stress conditions, it is important to match the carrier transport and failure mechanisms under operating conditions as closely as possible. Some research work cautions against performing drain-bias stress in the punch-through regime. Traditionally, punch-through has been characterized by a steep increase in matched source-drain current as the drain voltage is swept from 0 V to a stress voltage while gate, source and well are grounded. For further details, see A.14. However, with state-of-the-art extremely low-threshold-voltage devices, and short gate-lengths, source and drain currents may be matched over the entire range from 0 V to a stress voltage with a fairly constant slope. In such cases, it becomes difficult to distinguish the "punch-through" regime from the normal operating regime for the transistor. When performing drain-bias TDDB on such transistors using drain-bias stress with gate, source and well grounded, the measured time to breakdown is shorter than when the source connection is allowed to float or when the source-drain current is limited by using an offset technique. Further research is needed to determine whether or not drain-bias stress with gate, source and well grounded, can be used to accurately project time-to-breakdown under operating conditions.

In addition to three voltages and areas, measurements should also be performed at minimally two stress temperatures for the same voltage and area:

In most cases over a limited temperature range (e.g., 85 °C to 125 °C), the temperature dependence of t_{bd} follows an Arrhenius relationship to temperature $t_{bd\ 63\%} \propto exp[E_A/kT]$. Both the E- or V-model and the power-law model have a temperature dependence to the acceleration factor (see A.9).

4.3.2 Failure Detection for CVS

The device is considered to have failed when one of the following breakdown conditions has been detected:

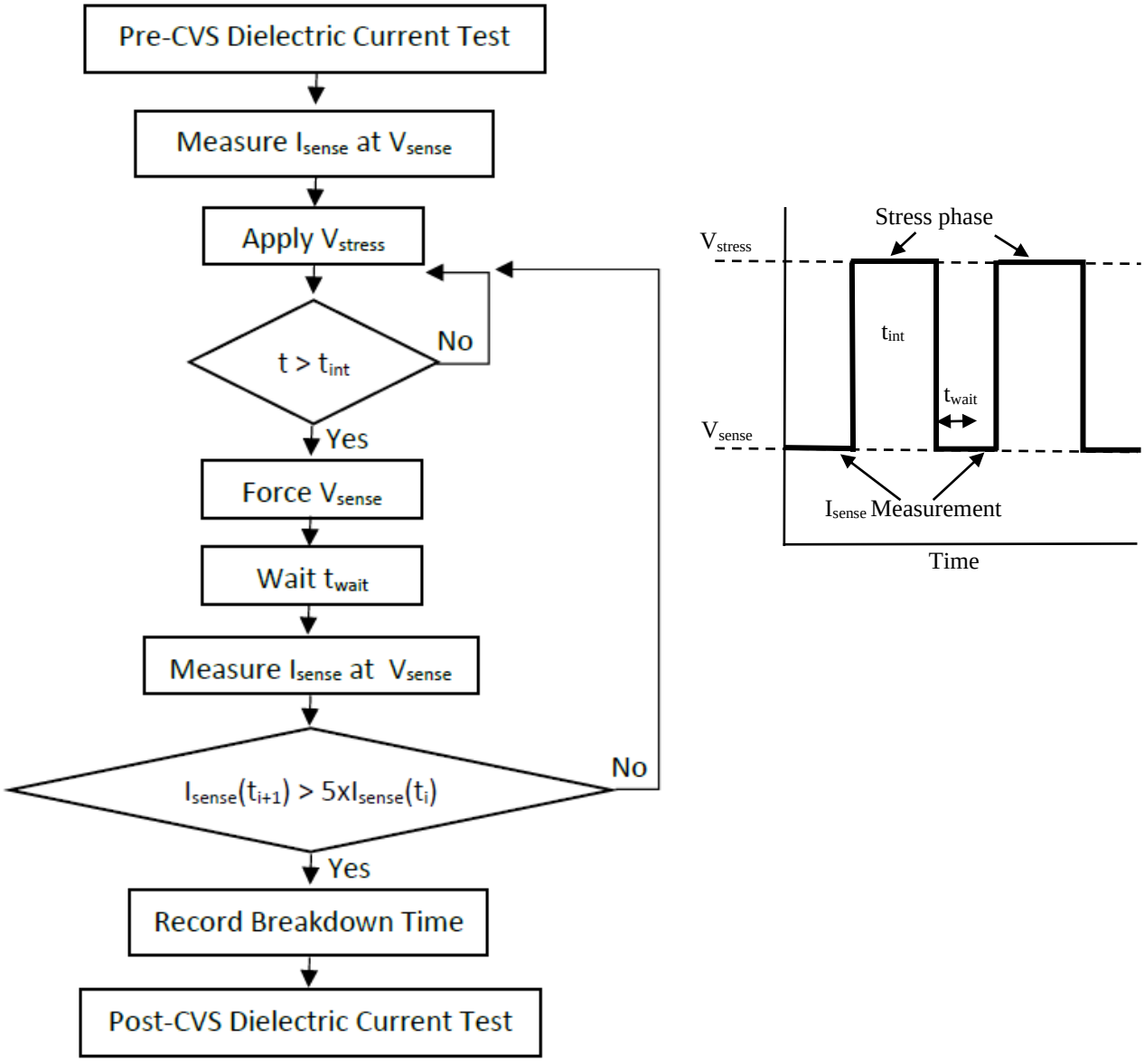
a) Increase in measured gate dielectric current at V_{stress}. For thicker dielectrics (th_{diel} >5 nm) or for small-area test structures, the dielectric often fails by a sudden increase (>10×) in measured dielectric current at V_{stress} ($I_{meas} > 10\times I_{previous}$). If this condition is met, the test is terminated and the post-test is performed. Note that the value of 10× increase is a recommended value. This value could range between 2× to 10× for actual breakdown events depending on test area, thickness, structure, or process.

b) Increase in current noise – see A.2.

c) Increase in current at sense voltage. This method monitors the increase of current at V_{sense} as a function of stress time to determine when soft-breakdown has occurred. In this technique at periodic time intervals (t_{int}) the voltage stress is interrupted and device current (I_{meas}) measured at low gate voltage (V_{sense}). After stress interruption and before the I_{meas} measurement, a wait time (t_{wait}) should occur to allow any transients to diminish which may occur in some test systems. The t_{int} value should be <1% of the anticipated t_{bd} while t_{wait} should be determined for each stress condition such that current settles to a constant value. Depending on the test system, typical t_{wait} values could be several seconds. I_{meas} should also be recorded to a data file. If $I_{meas}(t_{int} + 1) > r_{crit} \times I_{meas}(t_{int})$, then the device is defined as failed. The test is terminated and the post-test is conducted. Typical values of r_{crit} are between 2 and 10. The value depends on th_{diel}, A_{diel}, and V_{sense}. Figure 1 shows block and timing diagrams describing the stress interruption technique. It has been shown that periodic stress interruption does not affect the lifetime distributions for a variety of stress conditions.

d) Increase in current above I_{fail}, which can be set for V_{stress} or V_{sense}.

4.3.2 Failure Detection for CVS (cont'd)

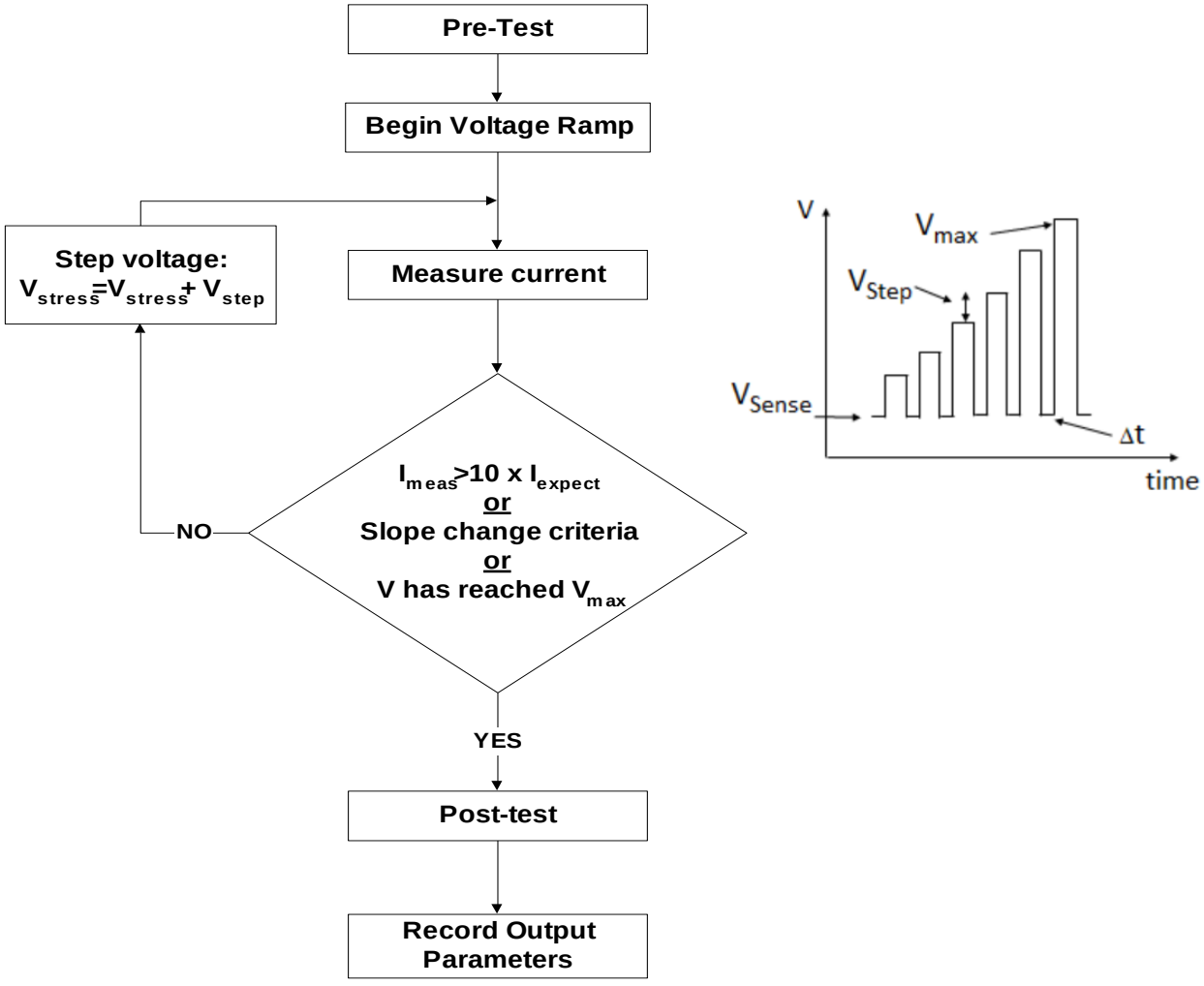


NOTE Block diagram and timing diagram showing the implementation of the stress interruption technique for monitoring the change in current at a given sense voltage (t_{int} should be <1% of the anticipated t_{bd}).

Figure 1 — Block and Timing Diagrams for the Stress Interruption Technique

4.4 Stress Component for V-Ramp

After the pre-screen, a linear or stepped voltage ramp is applied to the dielectric test structure. The voltage starts at V_{start} and ramps at a predefined ramp rate, or is stepped by the voltage V_{step} for duration t_{step}. During the voltage ramp the current is monitored at least as often as t_{step}. For the stepped voltage ramp, the current measurement should be delayed at each voltage step to allow displacement currents to settle. This implies that t_{step} must be longer than the instrument's settling time plus the measurement time of the test system. At each voltage step the measured dielectric current should be compared to the dielectric breakdown criterion. If the dielectric breakdown criterion is exceeded, the V-Ramp should be stopped. Figure 2 illustrates a typical voltage ramp test.



NOTE Basic voltage-ramp flow – inset shows an option to improve breakdown detection by dropping the voltage to V_{sense} between each of the ramped voltage steps.

Figure 2 — Basic Voltage Ramp Flow

A.3 describes a way to detect breakdown during V-Ramp by using a change in slope of the measured current.

4.4 Stress Component for V-Ramp (cont'd)

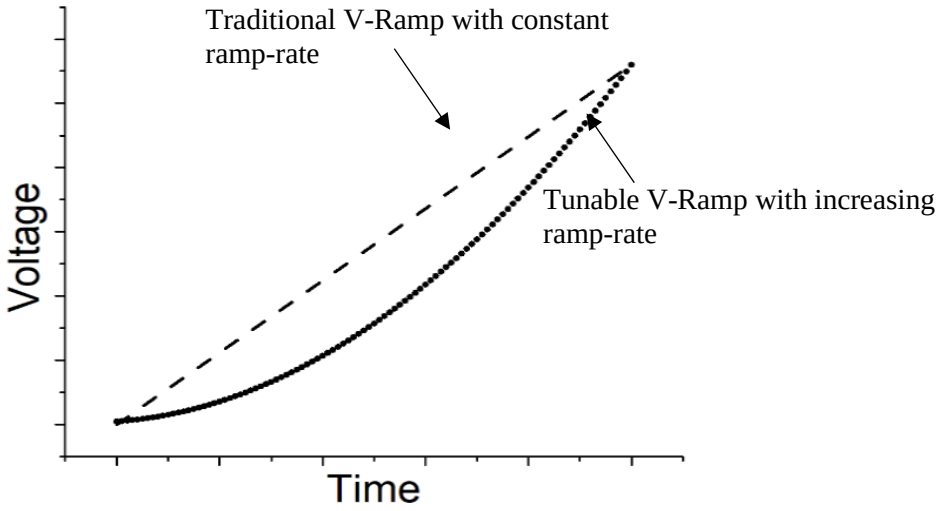
In some cases, breakdown can be more easily detected by dropping the voltage down to a sensing voltage V_{sense} after each step of the V-Ramp as shown in the inset of Figure 2. As in CVS, after stress interruption and before the I_{meas} measurement, a wait time (t_{wait}) should occur to allow any transients to diminish which may occur in some test systems. Setting V_{sense} at V_{use} allows the most sensitive breakdown detection capability.

Another variation on the V-Ramp is the Tunable V-Ramp. A lower ramp rate is used at the beginning of the V-Ramp in order to improve sensitivity to early failures and a higher ramp rate is used near the end of the V-Ramp stress. This is best accomplished by a continuously increasing ramp rate using the following formula for the voltage at the i^{th} step V_i:

$$V_i=(K*i + L)*(I + 1) + M \quad (4.3)$$

Where M is the initial voltage, L is the initial ramp rate, and K determines the level of ramp rate increase at each step. Figure 3 compares the traditional V-Ramp which uses a constant ramp-rate with the tunable V-Ramp where the ramp rate increases with time.

V-Ramp results can be converted to CVS results as described in A.4. However, periodic comparison to CVS results is necessary to confirm that the conversion process is continuing to give accurate results. The annex also describes how to derive the voltage acceleration factor using measurements with different ramp-rates.



NOTE Comparison of traditional V-Ramp (constant ramp-rate) with Tunable V-Ramp (ramp rate increases with time).

Figure 3 — Comparison of Traditional V-Ramp with Tunable V-Ramp

4.5 Test Procedure after Stress Component for CVS and V-Ramp

Following the stress component, a post-test similar to the pre-test should be conducted as follows:

1) Force V_{use}.
2) Measure I_{meas}. If $I_{meas} < I_{use}$ the device does not behave as expected and is considered to be passing the post-test.
3) Record I_{meas}.

4.6 Data Recording for CVS and V-Ramp

For all valid failures (see Table 2) record the following information as outlined here.

- t_{bd} - dielectric time-to-breakdown for CVS or V_{bd} – dielectric breakdown voltage for V-Ramp.
- I_{fail} - if the failure criterion is based on exceeding a given current at the stress voltage or at the sense voltage.
- V_{stress} for CVS.
- (Optional) V_{sense} if used.
- A_{diel_test}.
- T_{stress} junction temperature during stress.
- L_g and other relevant transistor characteristic for drain-bias stress.
- Failure category – the dielectric failure mode as defined in 4.7.

In addition it is recommended that the following parameters be recorded:

- I_{bd} (V_{stress}) - the dielectric breakdown current measured at the stress voltage.
- (Optional) I_{bd} (V_{sense}) - the dielectric breakdown current measured at V_{sense}.

4.7 Test Result Categories for CVS and V-Ramp

Because of the complexity of the dielectric test procedure, five possible test result types can occur. These types are listed in Table 2 and discussed in this section:

Table 2 — CVS and V-Ramp Dielectric Test Result Categories

Stress test result type	Pre-test	Stress component	Post-test	Recorded parameters
Type I	Fail	N/A	N/A	Test result type I, I_{meas}
Type II.x *	Pass	Fail	Fail	Test result type II, t_{bd} or V_{bd}, I_{fail} (if used), $I_{bd}(V_{stress})$ or $(\delta I_{meas})^2$ or $I_{bd}(V_{sense})$, I_{meas} from pre-test and post-test
Type III	Pass	Pass	Fail	Test result type III, I_{fail} (if used), I_{meas} from pre-test and post-test
Type IV	Pass	Fail	Pass	Test result type IV, t_{bd} or V_{bd}, I_{fail} (if used), $I_{bd}(V_{stress})$ or $(\delta I_{meas})^2$ or $I_{bd}(V_{sense})$, I_{meas} from pre-test and post-test
Type V	Pass	Pass	Pass	Test result type V, I_{meas} from pre-test and post-test

* The "x" indicates what breakdown criterion was used to detect breakdown: x = 1 (CVS: compliance or $I_{meas} > 10\times I_{previous}$, V-Ramp:compliance or $I_{meas} > I_{bd}$), x = 2 (CVS:current noise or Vramp:I-V slope increase of 3× (see A.3), x = 3 (CVS and V-Ramp:change in I_{meas} at V_{use}). Pass in the pre-test indicates that the initial dielectric current did not exceed the failure criteria, i.e., the dielectric was not "shorted". Pass in the post-test also indicates that the dielectric current did not exceed the failure criteria, i.e., it was not "shorted". Pass during the stress component indicates that dielectric breakdown was not detected with one of the failure criteria specified in 4.3.2 or 4.4.

4.8 Dielectric Time to Failure Prediction

After valid failures have been obtained (type II failures in Table 2) and acceleration parameters have been determined as described in 4.1 and 4.3.1, the time to reach a given failure fraction can be obtained. For gate dielectric thicknesses where direct-tunneling is the prevailing tunneling mechanism, the power-law model is accepted for voltage acceleration ($t_{bd\_63\%} \propto V^{-n}$), For gate dielectric thicknesses where Fowler-Nordheim tunneling is the prevailing tunneling mechanism, some research work confirms that the power-law model continues to be valid. Other research work provides evidence for the E-model (time-to-fail $\propto e^{-\gamma_E * E}$) in this thicker range of gate dielectric.

When benchmarking two processes using the E-model, one source of confusion can be the method used to calculate the electric field, E. However, more effective benchmarking can be performed by calculating the time to reach a certain failure fraction using the gate voltage instead of the field across the gate dielectric. Instead of using the E-model equation, $t_{bd\_63\%} \propto e^{-\gamma_E * E}$, it is common to use $t_{bd\_63\%} \propto e^{-\gamma_V * Vg}$.

4.8.1 Time to Reach a Given Failure Fraction Using Area and Failure-Fraction Scaling

As an example, the following equation shows how to use the power-law model to calculate the time to reach a given failure fraction due to gate dielectric breakdown in a given process:

$$t_{fail\_fraction} = \eta_{fail} * V^{-n} * exp(E_A/kT) * A^{-1/\beta fail} * (ln(1 - fraction)/ln(1 - 0.632))^{1/\beta fail} \quad (4.4)$$

In translating from failure of a test structure (tested at test voltage, V_{stress}, test junction temperature, T_{stress}, and test gate area, A_{diel_test}) to failure of a product (operated at use voltage, V_{use}, use junction temperature, T_{use}, and product gate area, $A_{diel\_product}$) due to gate dielectric breakdown, it is common to first determine $\eta_{fail\_product}$ using equation 4.5 and then go on to determine the time to reach a failure fraction required for a given product using equation 4.6.

$$\eta_{fail\_product} = \eta_{fail\_test} * (V_{use}/V_{stress})^{-n} * exp(E_A*(1/kT_{use} - 1/kT_{stress})) * (A_{diel\_product}/A_{diel\_test})^{-1/\beta fail} \quad (4.5)$$

$$t_{fail\_fraction\_product} = \eta_{fail\_product} * (ln(1-F_{required\_product})/ln(1 - 0.632))^{1/\beta fail} \quad (4.6)$$

For consumer electronics, some specifications require a failure fraction of <1/1000 (0.1% or 1000 ppm) while for automotive electronics, a specified failure fraction of <1/1000000 (1 ppm) is more common.

4.8.2 Time to Reach a Given Failure Fraction Using Failure-Fraction Scaling and Test Area

Although the method described below is generally valid, it is most commonly used for cases where progressive breakdown is observed. As mentioned in 4.1, the largest test area commonly used for current technologies is 1000 μm^2, and in many testing cases, much smaller areas are used. An alternative method to calculate the time to reach a given failure fraction for a product at a given voltage and temperature is to first use equation 4.2 to determine the product-level failure fraction based on a given test area, $F_{required\_product\_area}$.

4.8.2 Time to Reach a Given Failure Fraction Using Failure-Fraction Scaling and Test Area (cont'd)

A numerical example is shown using a test area of 1000 μm^2 and fail percentage requirement of 1 ppm.

Required failure fraction for the product (e.g. 1 ppm): $F_{required\_product}$ = 1/1000000
Test Dielectric Area: A_{diel_test} = 1000 μm^2
Product Dielectric Area: $A_{diel\_product}$ = 10^7 μm^2

$$ln\left(-ln\left(1 - F_{required\_product\_area}\right)\right) = ln\left(-ln\left(1 - F_{required\_product}\right)\right) - ln(A_{diel\_product}/A_{diel\_test}) \quad (4.7)$$

$$ln\left(-ln\left(1 - F_{required\_product\_area}\right)\right) = ln\left(-ln\left(1 - 10^{-6}\right)\right) - ln(10^7/10^3) \quad (4.8)$$

For this example, $F_{required\_product\_area}$ = 10^{-10}

Then calculate $\eta_{fail\_test\_area\_useVT}$ at the use voltage and temperature (after validating the test structures by observing that results from all test areas follow area-scaling as described by equation 4.2, η_{fail_test} for the largest of multiple test areas should be used in equation 4.9).

$$\eta_{fail\_test\_area\_useVT} = \eta_{fail\_test} * (V_{use}/V_{stress})^{-n} * exp(E_A*(1/kT_{use} - 1/kT_{stress})) \quad (4.9)$$

Then obtain the time to reach a given failure fraction for the product, using $F_{required\_product\_area}$ which combines the area-scaling from 4.5 and failure-fraction scaling from 4.6 into one term.

$$t_{fail\_fraction\_product} = \eta_{fail\_test\_area\_useVT} * (ln(1 - F_{required\_product\_area})/ln(1 - 0.632))^{1/\beta_{fail}} \quad (4.10)$$

Both equation 4.6 and 4.10 are valid ways to obtain the time to reach a given failure fraction for a product. In Annex A.8, the need for large sample size (~1000 datapoints) is demonstrated in order to accurately determine β_{fail} in cases where a single-sloped Weibull distribution is not obtained. As shown in Figure 9, multiple areas can be combined using equation 4.2 to create a distribution which covers a Weibit range similar to that covered by 1000 datapoints from a single area, and a fit to such a combined distribution can also be used to accurately determine β_{fail}.

4.8.3 Dielectric Time-to-failure Prediction Considerations

Since a given process always includes a small range of gate dielectric thicknesses, it is important to confirm that the time to reach a given failure fraction meets the required specifications even for thinner dielectrics in the range. A.11 describes this methodology.

It is useful to calculate the lifetime for cases where the stress is not applied constantly over time. Further details about dielectric breakdown lifetime under AC stress and ways to determine an approximate fraction of transistors in a circuit which may be under AC or DC stress at any point in time are discussed in A.10.

5 Current Stress for Dielectrics (J-Ramp and Bounded J-Ramp)

5.1 Pre-Stress Screen for J-Ramp and Bounded J-Ramp

1) Set and record stress temperature. Temperature should be controlled within ± 2 ^{o}C.
2) Determine typical I_{use} values at V_{use}.
3) Force V_{use}.
4) Measure I_{meas}.
5) If I_{meas} > 10×average value I_{use} determined from the pre-characterization test, record as initial failure.
6) Force I_{init}
7) Measure V_{meas}.
8) If $V_{meas} < V_{use}$, record as initial failure.

It is important to note that the measurement delay time between applying I_{init} and measuring V_{meas} must be long enough for the system to reach the correct equilibrium initial voltage. This may take several seconds for low forcing currents.

5.2 J-Ramp and Bounded J-Ramp Stress Component

Once the DUT passes the initial test, the J-Ramp test should start immediately. The initial stress current should be large enough to yield a short initial test time but small enough to detect early dielectric failures. (Typically, values of q_{bd} below 10^{-6} C/cm^2 are difficult to obtain due to system capacitance.) Initial stress current values often depend on test structure area, dielectric thickness and equipment capabilities.

The current should be stepped (step duration must be uniform) in logarithmic intervals such that the ratio of two successive steps is a constant factor, r_{step}, which must be less than sqrt(10) or 3.2. Caution should be exercised when comparing test data taken with different r_{step}-factors. During the ramp, the voltage across the dielectric is measured at evenly spaced intervals at least once per current step.

5.2 J-Ramp and Bounded J-Ramp Stress Component (cont'd)

Figure 4 displays the basic J-Ramp test flow diagram.

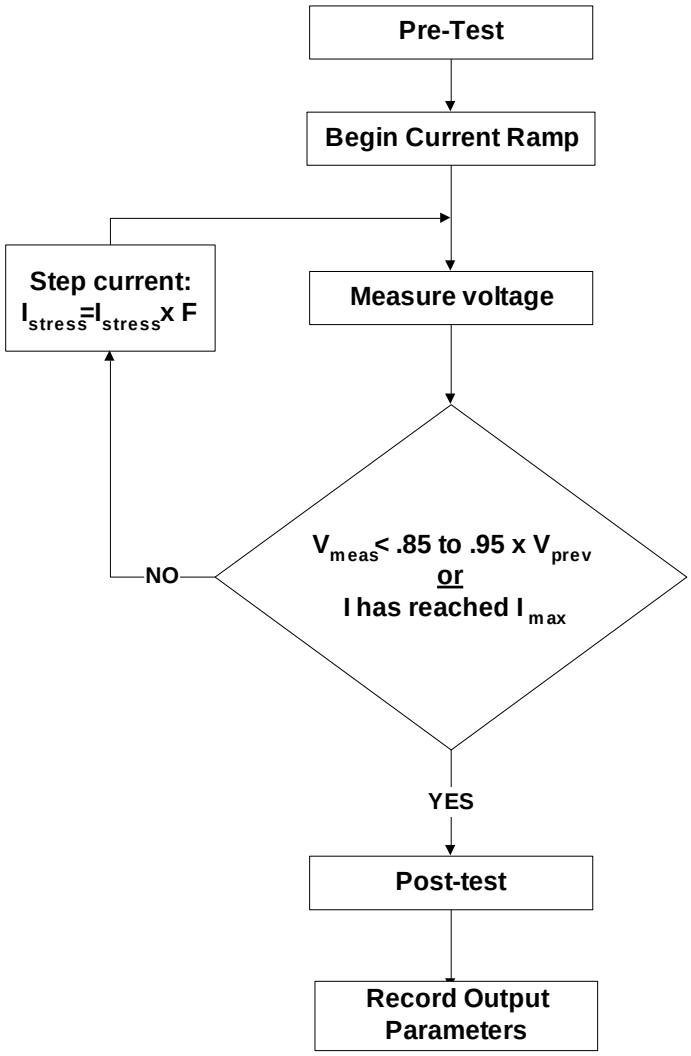


Figure 4 — Basic J-Ramp Flow Diagram

A variation of the J-Ramp test is the Bounded J-Ramp test. In the Bounded J-Ramp test the current is ramped to a constant value (prevents voltage overshoot) and is held at this value for the duration of the test. The current ramp should be performed in the same manner as the J-Ramp test. Once a specified current (I_{bound}) is reached, the current is held constant until breakdown. Empirical measurements have shown that lower current levels yield higher Q_{bd} and tighter measurement distributions. However, current levels that are too low result in test time that is too long.

The structure should be considered to have failed when the $V_{meas}(i) < 0.85*V_{meas}(i-1)$ where i denotes the i^{th} step in the ramp or the i^{th} measurement of voltage in a Bounded J-Ramp.

5.3 J-Ramp and Bounded J-Ramp Stress Test

1) Force I_{init}.

2) Measure V_{meas}.

3) If $V_{meas} > V_{use}$ the device has not undergone dielectric breakdown and should not be included in the dataset for analysis.

5.4 Dielectric Test Result Categories for J-Ramp and Bounded J-Ramp

Several parameters may be recorded as a result of the test including V_{bd} (maximum voltage attained during testing before breakdown), charge to-breakdown (Q_{bd}) obtained by integrating the current-time plot for all measurement intervals prior to breakdown.

Table 3 — Ramp and Bounded J-Ramp Dielectric Test Result Categories

Stress test result type	Pre-test	Stress component	Post-test	Recorded parameters
Type I	Fail	N/A	N/A	Test result type I, I_{meas}, V_{meas}
Type II	Pass	Fail	Fail	Test result type II, $V_{bd} = V_{prev}$, $q_{bd} = Q_{bd}/A_{diel}$, V_{meas} from pre-test and post-test
Type III	Pass	Pass	Fail	Test result type III, V_{meas} from pre-test and post-test
Type IV	Pass	Fail	Pass	Test result type IV, V_{meas} from pre-test and post-test
Type V	Pass	Pass	Pass	Test result type V, V_{meas} from pre-test and post-test

6 Bibliography

1. Degraeve, R. *et al.*, "On the Field Dependence of Intrinsic and Extrinsic Time-Dependent Dielectric Breakdown," *Proc. International Reliability Physics Symposium*, 1996, pp. 44-54.
2. Degraeve, R. *et al.*, "Reliability: a Possible Showstopper for Oxide Thickness Scaling," *Semicond. Sci. Tech.*, Vol. 15, 2000, p. 425.
3. Monsieur, F. *et al.*, "Determination of Dielectric Breakdown Weibull Distribution Parameters Confidence Bounds for Accurate Ultra-thin Oxide Reliability Predictions," *Proc. of ESREF*, 2001, p.1.
4. Snyder, E.S. and J. S. Suehle, "Detecting Breakdown in Ultra-Thin Dielectrics Using a Fast Voltage Ramp," *IEEE Integrated Reliability Workshop Final Report, Lake Tahoe, CA*, 1999, pp.118-123.
5. Brisbin, D. *et al*., "Influence of test techniques on soft breakdown detection in ultra-thin oxides," *Microelectronics Reliability*, 2002, Vol. 42, p. 35.
6. Pompl, T. *et al.,* "Influence of Soft Breakdown on NMOSFET Device Characteristics," *Proc. International Reliability Physics Symposium*, vol. 37, 1999, pp. 82-87.
7. Wang, B. *et al.,* "The Effect of Stress Interruption and Pulsed Biased Stress on Ultra-Thin Gate Dielectric Reliability," *IEEE Integrated Reliability Workshop Final Report, Lake Tahoe, CA*, 2000, pp. 74-79.
8. Richter, C.A. *et al.,* "A Comparison of Quantum-Mechanical Capacitance-Voltage Simulators," *IEEE Elec. Dev. Lett.*, vol. 22, 2001, pp. 35-37.
9. McPherson, J. and D.A. Baglee, "Acceleration Factors for Thin Gate Oxide Stressing," *Proc. International Reliability Physics Symposium*, Vol. 23, 1985, pp. 1-5.
10. McPherson, J.W. *et al.,* "Molecular Model for Intrinsic Time-Dependent Dielectric Breakdown in SiO2 Dielectrics and the Reliability Implications for Hyper Thin Gate Oxide," *Semicond. Sci. Technol*, vol. 15, 2000, pp. 462-470.
11. Hu, C. and Q. Lu, "A Unified Gate Oxide Reliability Model," *Proc. International Reliability Physics Symposium*, vol. 37, 1999, pp. 47-51.
12. Schuegraf, K.F. and C. Hu, "Metal-Oxide-Semiconductor Field-Effect-Transistor Structures Substrate Current During Fowler-Nordheim Tunneling Stress and Field Dioxide Reliability," *J. Appl. Phys.*, vol. 76, 1994, pp. 3695-3700.
13. Moazzami, R. *et al.,* "Temperature Acceleration of Time-dependent Dielectric Breakdown*," IEEE Trans. Elec. Dev.*, vol. 36, 1989, pp. 2462-2465.
14. Suehle, J.S. *et al.,* "Low Electric Field Breakdown of Thin SiO2 Films Under Static and Dynamic Stress," *IEEE Trans. Elec. Dev.*, vol. 44, 1997, pp. 801-808.
15. McPherson, J. *et al.,* "Comparison of E and 1/E TDDB Models for SiO2 under longterm/low-field test condition," *Tech. Digest IEDM*, 1998, pp. 171-174.
16. Hu, C. and Q. Lu, "A Unified Gate Oxide Reliability Model," *Proc. International Reliability Physics Symposium*, Vol. 37, 1999, pp. 47-51.
17. Cheung, K.P., "A Physics-Based Unified Gate-Oxide Breakdown Model," *Tech. Digest IEDM*, 1999, pp. 719-722.
18. Wu, E.Y. and J. Suñé, "Power-law Voltage Acceleration: A Key Element for Ultra-Thin Gate Oxide Reliability," *Microelectronics Reliability*, vol. 45, 1809-1834, 2005.
19. Hung, S.C. *et al.,* "Time-Efficient Characterization of Time-Dependent Gate Oxide Breakdown Using Tunable Ramp Voltage Stress (TRVS) Method for Automotive Applications," *Proc. International Reliability Physics Symposium*, 4B.3, 2021.
20. Wu, E.Y. and J. Suñé, "On Voltage Acceleration Models of Time to Breakdown—Part I: Experimental and Analysis Methodologies," in *IEEE Trans. Elec. Dev.*, vol. 56, no. 7, pp. 1433-1441, July 2009.

6 Bibliography (cont'd)

21. Aal, Andreas, "Comparison between Gate Oxide Lifetime Models with *R*series and Trapping Effect Correction in the FN-Regime," *IEEE Integrated Reliability Workshop Final Report, Lake Tahoe, CA*, 2012, pp. 99-104.
22. Kuo, Ren-Jay *et al.,* "Antifuse OTP Cell in a Cross-Point Array by Advanced CMOS FinFET Process," in *IEEE Trans. Elec. Dev.*, vol. 66, no. 4, pp. 1729-1733, April 2019.
23. Kaczer, B. *et al.,* "Impact of MOSFET Gate Oxide Breakdown on Digital Circuit Operation and Reliability," *IEEE Trans. Elec. Dev.*, vol. 49, no. 3, pp. 500-506, March 2002.
24. Kaczer, B. *et al.*, "Experimental Verification of SRAM Cell Functionality after Hard and Soft Gate Oxide Breakdowns," *Conference on European Solid-State Device Research*, 2003, pp. 75-78.
25. Rodriguez, R. *et al.,* "The Impact of Gate-Oxide Breakdown on SRAM Stability," *IEEE Elec. Dev. Lett.*, vol. 23, no. 9, pp. 559-561, September 2003.
26. Wu, E. *et al.*, "A Viable and Comprehensive TDDB Assessment Methodology for Investigation of SRAM Vmin Failure," *Tech. Digest IEDM,* 2009, pp. 397-400.
27. Mueller, K. *et al.*, "6-T Cell Circuit Dependent GOX SBD Model for Accurate Prediction of Observed VCCMIN Test Voltage Dependency," *Proc. International Reliability Physics Symposium,* 2004, pp. 426-429.
28. Wu, Ernest Y. *et al.*, "On the Weibull Shape Factor of Intrinsic Breakdown of Dielectric Films and its Accurate Experimental Determination – Part I: Theory, Methodology, Experimental techniques," *IEEE Trans. Elec. Dev.*, vol. 49, no. 12, pp. 2131-40, December 2002.
29. Wu, Ernest *et al.*, "Off-State Mode TDDB Reliability for Ultra-Thin Gate Oxides: New Methodology and The Impact of Oxide Thickness Scaling," *Proc. International Reliability Physics Symposium,* 2004, pp. 84-94.
30. Chen, I.K. *et al.*, "The Physical Mechanism Investigation of Off-State Drain Bias TDDB and its implication in Advance HK/MG FinFETs," *Proc. International Reliability Physics Symposium*, 4A.2, 2018.
31. M. Toledano-Luque *et al.*, "Off-state TDDB in FinFET Technology and its Implication for Safe Operating Area," *Proc. International Reliability Physics Symposium, 2A.3*, 2021, pp. 1-6.
32. Nminibapiel, D. *et al.,* "Method to evaluate off-state breakdown in scaled Tri-gate technologies," *Proc. International Reliability Physics Symposium*, 9B.1, 2022.
33. Liu, Wen *et al.*, "Robust Off-State TDDB Reliability of n-LDMOS" *Proc. International Reliability Physics Symposium*, P26, 2022.
34. Joshi, Kaustubh *et al*., "A Statistical Learning Model for Accurate Prediction of Time-Dependent Dielectric Degradation for Low Failure Rates," *Proc. International Reliability Physics Symposium*, 4E.4, 2019.
35. Furukawa, T. *et al*., "Accelerated gate-oxide breakdown in mixed-voltage I/O circuits," *Proc. International Reliability Physics Symposium*, 1997, pp. 169-173.
36. Vandemaele, M. *et al*., "Simulation Comparison of Hot-Carrier Degradation in Nanowire, Nanosheet and Forksheet FETs," *Proc. International Reliability Physics Symposium*, 6A.2, 2022, pp. 1-9.
37. Kaczer, B. *et al.,* "Observation of hot-carrier-induced nFET gate-oxide breakdown in dynamically stressed CMOS circuits," *Tech. Digest IEDM*, 2002, pp. 171-174
38. Garba-Seybou, Tidjani *et al.*, "Location of Oxide Breakdown Events under Off-state TDDB in 28nm N-MOSFETs dedicated to RF applications," *Proc. International Reliability Physics Symposium*, 2023, 4C.2 pp. 1-8.
39. Joshi, K. *et al.*, "A Detailed Comparison of Various Off-State Breakdown Methodologies for Scaled Tri-Gate Technologies." *Proc. International Reliability Physics Symposium*, 2023. 9A.2 pp. 1-6.
40. Hai, J. *et al.,* "Integrated Test Circuit for Off-State Dynamic Drain Stress Evaluation." *Proc. International Reliability Physics Symposium*, 2023. 4B.4 pp. 1-6.

Annex A (Informative) Test Structure and Data Analysis

A.1 Test Structure Overview

Test Structures may be capacitors or transistors. Typically, capacitors in an integrated circuit use dielectrics thicker than 10 nm, while transistor gate dielectrics are thinner than 10 nm. This section begins with plasma charging effects which are important to avoid for all test structures and then gives more specific information about test structures for dielectrics thinner than or thicker than 10 nm.

A.1.1 Avoiding Design Pitfalls

Good test structure design must minimize several dielectric-current-related effects. The test structure must be able to sustain high stress currents. The design must maintain a uniform electric field over the dielectric area in spite of potential parasitic voltage drops in the connecting electrodes associated with the high stress current. The performance of the test structure must not degrade due to high-current-induced interconnect, contact, or via electromigration. Large peak currents may occur during the dielectric rupture event causing metal open circuits or contact burnout. Displacement currents occurring during the stress must be differentiated from the actual stress current. Displacement currents are due to the charging of parasitic instrument cable and switching matrix capacitance as well as the test structure itself. The parasitic instrumentation capacitance can be greater than the actual test structure capacitance. Scribe-line test structures may receive slightly different processing conditions compared to dielectrics within the circuit. This may lead to TDDB test result differences.

A.1.1.1 High Currents During Dielectric Breakdown

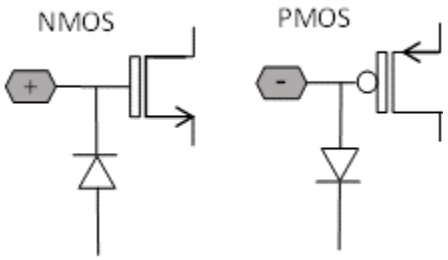
During the breakdown event, extremely high transient currents flow. Initially the discharge current comes from the charge stored on the test structure. The peak current is limited by the current dynamics in the breakdown plasma (at the point of breakdown, the current becomes so high that melting can occur and a localized plasma can form), and by the resistance of the connecting electrodes. The greater the local dielectric voltage, the greater the energy released into the breakdown event. If the test structure is an array of many small transistors, the whole array charge is conducted through the gate or drain connection to the transistor that ruptured. If the interconnect is narrow or has few contacts, the connection may open and the breakdown event may go undetected. The greater the capacitance voltage, the greater the energy provided by the instrumentation cables and supplies for the breakdown event. When the test structure’s voltage suddenly drops, the instrumentation cables behave as a transmission line. Extremely high peak current occurs, limited mainly by the cable’s inductance and characteristic dynamic impedance. The series resistance of the test structure may also limit the current, but since the series resistance is designed to be low, high current can flow. This high current flows through the interconnect and can cause via and contact open circuit damage. Increasing interconnect width and number of vias and contacts can prevent this damage from occurring.

A.1.2 Process Plasma Charging Effects

Test structures with a variety of charge collection antennas are used to assess the effect of plasma charging damage on dielectric reliability [1]. The effect is related to the efficiency of charge collectors called antennas. Test structures often have large bond pads and a different ratio of contact to test area than transistors in a product. Because of these differences the test structure may not be representative of product transistors. Therefore test structure evaluation using different antenna types and sizes is desirable. The gate-electrode antenna ratio (charge collection efficiency) is defined as a ratio of the antenna area to the thin dielectric area. It can also be defined in terms of the ratio of the antenna perimeter to the capacitor area. Either factor can be important depending on the processing conditions. Test structures with antennas that have different area and edge components allow processing steps and their effect on dielectric quality to be characterized. In addition, metal interconnect and via processing including the bond pad can cause dielectric charging damage. Metal interconnect charging damage can be characterized by metal antennas with different interconnect areas and edge ratios. The via or contact process charging damage can be characterized by antennas with different number of contacts or vias. In this case the number of contacts or vias divided by the capacitor area is a measure of the charge collection efficiency. Key antenna rules are implemented in Design Manuals to prevent impact to the product of possible plasma-induced damage during processing.

The goal of this standard is to evaluate the intrinsic gate dielectric breakdown of a given MOSFET as part of a full build not affected by uncontrolled process events such as plasma induced damage. In order to avoid/reduce charge-collecting antenna effects, the test structure should be separated from the test pad area during the most relevant plasma steps (deposition or etching). This can be achieved by the following two approaches:

1. Using transient fuses [2] in the lowest metal interconnect scheme and combined transient fuse and bridging [3] techniques at the highest metal level to shunt charge accumulation from sensitive gate areas.
2. The addition of protection diodes at the test pad/gate may also be appropriate for reducing charge collection [4]. Figure 5 shows the typical configuration for gate-bias TDDB inversion testing of an NMOS and a PMOS. With the addition of a diode, the gate current measured during gate-bias TDDB will include the reverse-bias leakage of the diode. In most cases, the source and drain currents can be summed to determine the transistor gate current. If accumulation testing is performed, specific test structures will need to be designed with the diodes flipped.



NOTE NMOS and PMOS test structures with protection diodes to reduce plasma damage during processing. These are designed for gate-bias inversion testing (NMOS Vg = +bias, PMOS Vg = -bias, as shown) or drain-bias testing (NMOS Vd = +bias, PMOS Vd = -bias) with all other terminals grounded.

Figure 5 — NMOS and PMOS Test Structures with Protection Diodes

A.1.3 Test Structures for Dielectrics Thinner than 10 nm

Relatively small area test structure should be used in order to accurately detect t_{bd} or V_{bd} and to minimize extrinsic defects and determine the intrinsic dielectric breakdown behavior. The failure time is a function of area, even in the absence of gross defects. Background leakage current for dielectrics less than 5 nm can become significant for large area test structures. Device breakdown will become very difficult to detect if the jump in current associated with the breakdown event is of the same magnitude as the background leakage current. In addition, the series resistance associated with larger device area can cause a significant voltage-drop so that a constant voltage is not maintained along the channel of the MOSFET, or a constant voltage is not maintained throughout the area of a capacitor test structure. For dielectrics with th_{diel} less than 2 nm it may become necessary to limit the test structure area to ~100 μm^2 or smaller depending on the magnitude of the background current**.** Transistor arrays are most commonly used. It is important to minimize the series resistance of the interconnects to the test structure. Series resistance effects cause parasitic voltage drops reducing the effective voltage and electric field across the dielectric. In some cases, structures may be chosen to emphasize specific geometry, e.g., poly edge or field oxide isolation edge. Test structures with large area polysilicon or metal electrodes (not necessarily large antenna ratios) are susceptible to plasma charging and damage. This should be considered in assessing TDDB data. Diode protection may be appropriate, although it restricts the applied stress voltage polarity and maximum voltage that can be used. Therefore, antenna structures should also be characterized to assess plasma charging.

A.1.4 Test Capacitors for Dielectrics Thicker than 10 nm

Three basic types of test capacitors are essential for dielectric reliability characterization. These are area intensive, gate-edge intensive, and isolation-edge intensive capacitors. TDDB testing is commonly used to measure intrinsic dielectric wearout. Wearout is often associated with area intensive capacitors. However it is important to recognize that product lifetime, even intrinsic wearout, may be dominated by the dielectric reliability at the gate edge or at the edge of the isolation dielectric. An area-intensive capacitor is one with a large area component relative to edge components. A gate-edge capacitor is designed with maximum gate edge over thin dielectric and has minimum area and isolation edge. Similarly, the isolation-edge capacitor maximizes the isolation edge component while minimizing the other two. With three capacitors and three different ratios of the three components it is possible to determine each failure density contribution to the overall dielectric reliability.

The test structures should be designed to keep the electric field uniform over the capacitor area. One way to accomplish this is to use rectangular test structures. The gate electrode should have multiple contacts along one of the long edges. Substrate contacts should be placed along the long edge opposite the gate contacts. Due to high stress currents there will be voltage drops in the gate electrode and substrate. By placing the contacts along opposite sides of the capacitor the voltage drops will compensate one another, keeping the electric field as uniform as possible over the capacitor area. The typical aspect ratio of rectangular capacitors is 1:5 to 1:10. The greater the sheet resistance difference between the gate and substrate, the larger the aspect ratio should be. To avoid high stress current and possible electromigration, use wide metal and polysilicon interconnects. The number of contacts and vias should be sufficient to avoid contact electromigration. Although the maximum number and thickness of conductor lines is desirable for minimal resistive drops, the area above the capacitors should be kept as clear as possible to facilitate failure analysis. The test structure total series resistance should be as low as possible. Series resistance effects cause parasitic voltage drops leading to reduced stress voltages and currents and longer TDDB lifetime than expected.

A.1.4 Test Capacitors for Dielectrics Thicker than 10 nm (cont'd)

Metal gate capacitors are prone to self-healing. Self-healing occurs when the metal vaporizes above the rupture site creating an open rather than a dielectric short. Breakdown events of this type may go undetected. Substrate voltage drops can alter the stress conditions of nearby capacitors when testing devices in parallel. Separate wells and well contacts are recommended for each device. Adding substrate guard rings can help control the substrate potential. The substrate potential needs to be considered at normal stress conditions and after capacitor breakdown.

If lifetime tests are being performed at high temperature, the sheet resistance increases. This increased sheet resistance may affect the electric field uniformity across the capacitance area and the test structure series resistance.

References for A.1:

1. As this standard is being balloted, a JEDEC task group (142_3) is creating a plasma damage standard.
2. S. Krishnan *et al.*, "A Transient Fuse Scheme for Plasma Etch Damage Detection," *3rd International Symposium on Plasma Process-Induced Damage*, 1998, pp. 201-204.
3. P. Simon *et al.*, "Antenna Ratio Definition for VLSI Circuits [Plasma Etch Damage]," *4th International Symposium on Plasma Process-Induced Damage*, 1999, pp.16-20.
4. A. Martin, "Plasma induced Charging Damage: From Appropriate MOS Test Structures to Antenna Design Rules, A Comprehensive Process Qualification Procedure, "*IEEE International Integrated Reliability Workshop (IIRW)*, 2020, pp. 1-8.

A.2 Breakdown Detection for CVS by Increase in Current Noise

At a soft-breakdown dielectric event, the measurement noise increases. This increase in noise can be detected by analyzing the current measurement data using variance techniques. This test description assumes that the test system noise has already been determined as described in the pre-characterization test (See 4.2). In this test, six consecutive current values of $I_{meas}\ (i)$ to $I_{meas}\ (i+5)$ are recorded and the current noise $(\delta I_{meas})^2$ is calculated from these values as:

$$(\delta I_{meas})^2 = \frac{\sum_{i=1}^{i=5}\{I_{meas}(i)\}^2 - \frac{[\sum_{i=1}^{i=5} I_{meas(i)}]^2}{5}}{4} \qquad \text{(A.2.1)}$$

The current noise is continuously calculated by adding a new current value and deleting the first value in the six-point set (i.e., a sliding sample set: $I_{meas}\ (i+1)$ to $I_{meas}\ (i+5)$). If the current noise increases by 500× over the baseline value for at least five additional calculations, then the device is defined as having failed. The additional calculations performed past the detection of breakdown assures that the noise increase is sustained and not a result of a random fluctuation or a transient noise increase. The test is then terminated and the post-test performed. Note that value of 500× increase is a recommended value. This value could range between 200× and 500× for actual soft breakdown events depending on capacitor area, thickness, structure, or process.

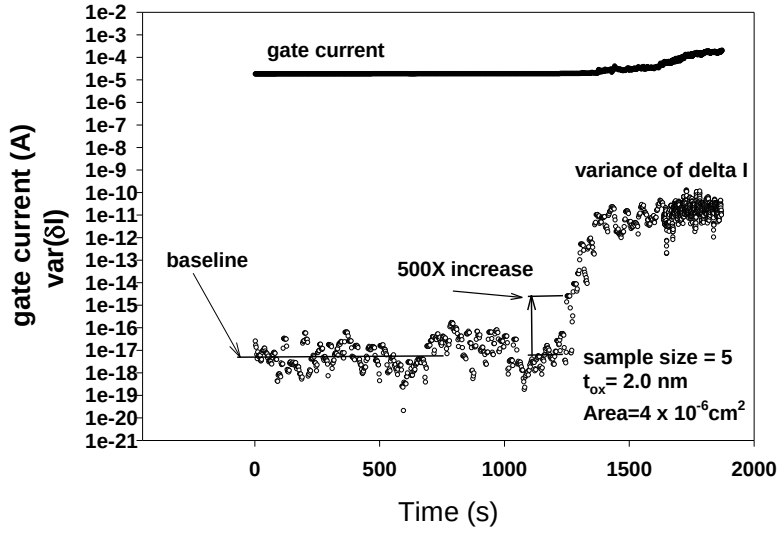
A.2 Breakdown Detection for CVS by Increase in Current Noise (cont'd)

It may be desirable to compensate for increasing values of I_{meas} during the stress due to trapping or stress-induced leakage current. In this case, the value of $(\delta I_{meas})^2$ can be calculated from the variance of five values of the *difference* between the $I_{meas}(i+1)$ - $I_{meas}(i)$ data points in the six point sample as follows:

$$(\delta I_{meas})^2 = \frac{\sum_{i=1}^{i=5}\{I_{meas}(i+1)-I_{meas}(i)\}^2 - \frac{[\sum_{i=1}^{i=5} I_{meas}(i+1)-I_{meas}(i)]^2}{5}}{4} \qquad \text{(A.2.2)}$$

Figure 6 illustrates a typical example of implementing the variance method for detecting breakdown. The example is for a 2.0 nm thick SiO_2 sample with an area of 4 x 10^{-6} cm^2.

Note that more than a four-order-of-magnitude increase in the current noise is observed at the onset of dielectric breakdown. Ultra-thin gate dielectrics have been observed to exhibit rapid current transients and random telegraph signals (RTS) before breakdown. Care must be taken to avoid detecting breakdown under these conditions. A technique described in [1] has been shown to reduce sensitivity to RTS and other transient behavior. Valuable information can also be found in [2].



NOTE Stress gate current vs. time and gate current noise vs. time for a 2.0 nm thick SiO_2 film with an area of 4×10^{-6} cm^2. The onset of breakdown is detected by a >500× increase in the current noise.

Figure 6 — Stress Gate Current and Gate Current Noise

References for A.2:

1. P. Roussel *et al.*, "Accurate and Robust Noise-Based Trigger Algorithm for Soft Breakdown Detection in Ultra Thin Oxides," *Proc. International Reliability Physics Symposium*, Vol. 39, 2001, pp. 386-391.
2. G. B. Alers *et al.*, "J-Ramp on Sub-3 nm Dielectrics: Noise as a Breakdown Criterion," *Proc. International Reliability Physics Symposium*, Vol. 37, 1999, pp. 410-413.

A.3 Breakdown Detection for V-Ramp using a Change in Slope

Another possible criterion specifies that dielectric breakdown occurs when the logarithmic slope of I_{meas} vs. V_{stress} curve increases by a factor greater than the previously calculated slope. Typical exit values are an increase in slope that is greater than 2.5× to 5× the previous slope value with 3× being the recommended value. Measurement noise restricts use of this failure criterion to dielectric currents at least 10× the noise floor of the instrument. The previous logarithmic slope ($Slope_{prev}$) is calculated from

$$Slope_{prev} = \frac{abs(ln\left(abs\left(I(i-1)\right)\right) - ln\left(abs\left(I(i-2)\right)\right))}{V(i-1)-V(i-2)} \quad \text{(A.3.1)}$$

where *I(i-1)*, *V(i-1)* and *I(i-2)*, *V(i-2)* are the measured currents and voltages of the previous two data points and "*abs*" is the absolute value. The most recent slope ($Slope_{new}$) is calculated

$$Slope_{new} = \frac{abs(ln\left(abs\left(I(i)\right)\right) - ln\left(abs\left(I(i-1)\right)\right))}{V(i)-V(i-1)} \quad \text{(A.3.2)}$$

where *I(i)*, *V(i)* and *I(i-1)*, *V(i-1)* are the measured currents and voltages of the most recent and previous data points, respectively. For example, a dielectric failure is detected if the most recent measured slope ($Slope_{new}$) is 3× greater the previous calculated slope ($Slope_{prev}$).

A.4 Ramp and CVS Comparison

The Weibull shape factor can be determined using V-Ramp data

As described in 4.3.1, $t_{bd\_63\%}$ follows a power-law relationship with stress voltage ($t_{bd\_63\%} \propto V^{-n}$) where *n* is the power-law exponent. Data from V-Ramp stress with at least three different linear ramp rates can also be used to determine the value *n*, as follows:

$$Ramp\text{-}Rate \propto V_{bd}^{\,n+1} \quad \text{(A.3.3)}$$

When *n* has been determined, the Weibull shape factor, β, can also be determined by plotting V_{bd} on a Weibull scale. The slope of V_{bd} plotted on a Weibull scale will be $\beta \times (n+1)$. For reference, CVS data follows this relationship

$$ln(-ln(1-F)) = \beta ln\left(\frac{t_{bd}}{t_{bd\_63\%}}\right) \quad \text{(A.3.4)}$$

V-Ramp data follows this relationship

$$ln(-ln(1-F)) = \beta(n+1)ln\left(\frac{V_{bd}}{V_{bd\_63\%}}\right) \quad \text{(A.3.5)}$$

A.4 Ramp and CVS Comparison (cont'd)

Although the power-law model was discussed above, CVS and V-Ramp can also be correlated based on the exponential model. While the following holds true for the power-law model,

$$t_{bd} = \frac{V_{ref\_cvs}}{RR \cdot (n+1)} \left(\frac{V_{bd}}{V_{ref\_cvs}} \right)^{n+1} \quad \text{(A.3.6)}$$

where RR is the ramp-rate and V_{ref_cvs} is the voltage used for the constant-voltage stress, the following is true for the exponential model where γ_V is the voltage acceleration factor in the exponential model.

$$t_{bd} = \frac{\exp\left(\gamma_V (V_{bd} - V_{ref\_cvs})\right)}{\gamma_V \cdot RR} \quad \text{(A.3.7)}$$

More details and derivations can be found in [1, 4]. Historical details about correlation between CVS and V-Ramp can be found in [2, 3, 5-8].

This method is based on the idea that damage which accumulates during the V-Ramp is equivalent to damage which accumulates during CVS, so this conversion between CVS and V-ramp can be used for extrinsic breakdown as well as intrinsic breakdown.

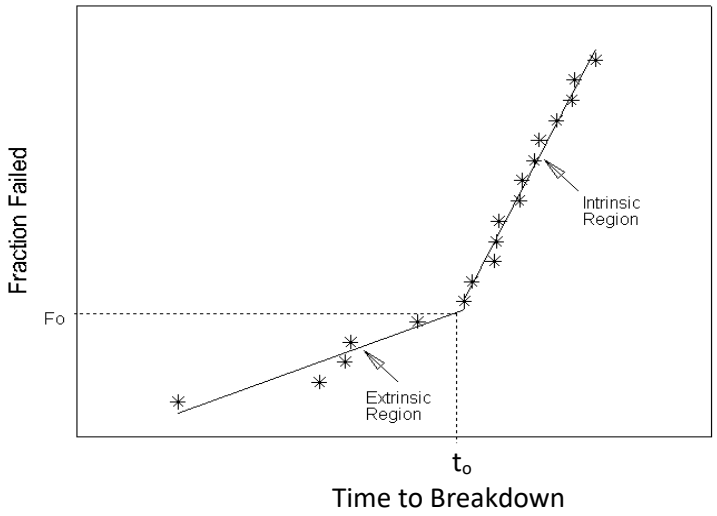
References for A.4:

1. A. Kerber *et al.,* "Reliability Screening of High-k Dielectrics Based on Voltage Ramp Stress," *Microelectronics Reliability* Vol. 47, 2007, pp. 513-517
2. A. Berman *et al.,* "Time-Zero Dielectric Reliability Test by a Ramp Method," *Proc. International Reliability Physics Symposium* 1981, pp.204-209.
3. S.C. Fan *et al.,* "Accurate Characterization on Intrinsic Gate Oxide Reliability Using Voltage Ramp Tests," *Proc. International Reliability Physics Symposium* 2006, pp.625-626.
4. A. Aal *et al.,* "A Comparison Between V-Ramp TDDB Techniques for Reliability Evaluation," *IEEE International Integrated Reliability Workshop Final Report* 2008, pp.133-136.
5. B. Kaczer *et al.,* "Gate Oxide Breakdown in FET Devices and Circuits: From Nanoscale Physics to System-Level Reliability," *Microelectronic Reliability* 47, 2007, pp. 559-566.
6. A. Aal *et al.,* "TDDB Data Generation for Fast Lifetime Projections Based on V-Ramp Stress Data," *IEEE Transactions on Device and Materials Reliability* 7, 2007, pp.278-284.
7. A. Aal *et al.,* "Fast Prediction of Gate Oxide Reliability – Application of the Cumulative Damage Principle for Transforming V-ramp Breakdown Distribution into TDDB Failure Distribution," *IEEE International Integrated Reliability Workshop Final Report*, 2006, pp.182-185.
8. S.C. Hung *et al.,* Time-Efficient Characterization of Time-Dependent Gate Oxide Breakdown Using Tunable Ramp Voltage Stress (TRVS) Method for Automotive Applications, *Proc. International Reliability Physics Symposium*, 4B.3, 2021.

A.5 Data Analysis

The use of the procedure about to be described is recommended when only catastrophic and initial failures are found during the tests, or when other types of failures are found and can be shown to resemble catastrophic failures. Refer to 4.7 and 5.4 for the definitions of the failure categories. The additional failure types referenced in the above sections are typically due to some artifact of the structure or the test that makes the data suspect. For example, in the V-Ramp [1] test, masked catastrophic failures may be due to an insufficiently high current compliance of the power supply or to excessive series resistance of the structure. Similarly, non-catastrophic failures may result from dielectric healing, and the "other" types of failures could be caused by poor probe-to-pad contacts or to open circuits elsewhere in the tester or structure. It should be remarked that ideally only catastrophic failures should be obtained. The presence of other failure categories suggests that either defect elimination work or test modification is needed. The data analysis begins by plotting the cumulative breakdown distribution on a Weibull or probability scale versus V_{bd} or t_{bd} . For thicker dielectrics with large test structure area, such a plot often results in a bimodal or other multi-modal distribution. Figure 7 shows a typical bimodal distribution, displaying an extrinsic and an intrinsic population. Analysis techniques are available that allow the separation of this distribution into two single mode distributions, namely the intrinsic and the extrinsic distributions [2]. Other examples of bimodal distributions can be found in [3]. The analysis of the breakdown charge density data is very similar to the analysis of the voltage breakdown or time to breakdown data. The principal difference is that the breakdown charge density is plotted on a logarithmic scale. This scale is necessary because the breakdown charge density typically ranges over several orders of magnitude. The vertical axis still has a normal probability scale as before. Usually, the resulting distribution is bimodal, and can be interpreted in terms of an intrinsic and an extrinsic population. The same techniques as before [2] can be used to separate the distributions for these two populations. Typical values of breakdown charge density for dielectrics in the 10 nm to 30 nm range are from 0.1 C/cm^2 to 10 C/cm^2.

Three parameters are of significant interest in analyzing these plots: the median and the standard deviation of the intrinsic population, and the percent of the population that is defective. The parameters pertaining to the intrinsic population are readily obtained after it is separated from the extrinsic population, as described in the previous sections. The percent defective is given by F_0 in Figure 7.



NOTE t_0 is the time that separates the intrinsic from the extrinsic distribution. F_0 is the fraction defective.

Figure 7 — Cumulative Breakdown Distribution Versus Time to Breakdown

A.5 Data Analysis (cont'd)

Notice that F_0 is not a direct measure of defect density because F_0 depends on the size of the test structure, as discussed further in Annex B. On the other hand, the intrinsic distribution is independent of test structure size. This property can be used to verify that a particular population is intrinsic, simply by testing structures with different gate dielectric areas.

The ideal plot of breakdown data (estimated breakdown electric field or breakdown charge density) should have no extrinsic population (no "defective" dielectrics), a very small main population standard deviation and a very large median. This set of conditions implies that there would be no weak dielectrics to fail early, and that the intrinsic dielectrics would not fail until well beyond the intended device life. The most effective way to use these data is to continuously drive the distribution closer to its ideal shape by instituting appropriate process controls. A secondary use is to determine the dielectric defect density, although the extremely large quantity of test data required to measure typical defect densities often makes this impracticable. Data output from the wafer-level dielectric testing discussed in this procedure has many potential uses. The most beneficial use of these data in the manufacturing environment is for process control and reliability trend assessment. By using the results and analysis of these tests in a tight loop process improvement system, the manufacturer can continually monitor and improve the reliability of the product. In addition to serving as a monitor function, it is intended that these tests be inserted into a manufacturer's overall process nodes through extensive design-of-experiments activities. Reliability trend assessment may be accomplished using standard Statistical Process Control procedures [4, 5]. Because Q_{bd} depends on area, design, ramp rate, step size and reading frequency, it should only be used as a process control monitor, not for determining relative dielectric quality for different structures and/or measurement techniques. Any comparisons should be performed using Bounded J-Ramp on the identically designed structure with the same dielectric thickness. Naturally, decreasing the step size and increasing the measurement frequency can increase measurement accuracy.

References for A.5:

1. ASTM F-17-71 *Test Method for Evaluating Gate Oxide Integrity by Voltage Ramp Technique*, American Society for Testing and Materials, West Conshohocken, PA, 1998.
2. D. S. Peck and O. D. Trapp, *Accelerated Testing Handbook*, Technology Associates, 1981.
3. R. Degraeve *et al.*, "On the Field Dependence of Intrinsic and Extrinsic Time-Dependent Dielectric Breakdown," *Proc. International Reliability Physics Symposium*, 1996, pp. 44-54.
4. K. Ishikawa, *Guide to Quality Control*, Asian Productivity Organization, 1982.
5. EIA/JEDEC Standard 19, *General Standard for Statistical Process Control (SPC)*, Electronic Industries Association, Washington D.C. 1989.

A.6 Determining Failure Rates

Given that a device under test survives up until time t, the instantaneous failure rate at time t is given by the hazard function *h(t)* as:

$$h(t) = \frac{f(t)}{1 - F(t)} \tag{A.6.1}$$

where *f(t)* is the probability density function (PDF) and *F(t*) is the cumulative distribution function (CDF) or unreliability function. The average failure rate of survivors over the time interval t_1 to t_2 is the integration of the hazard function *h(t)*, i.e.,

$$AFR(t_1, t_2) = \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} h(t')dt' = \frac{-\ln(1 - F(t_2)) + \ln(1 - F(t_1))}{t_2 - t_1} \tag{A.6.2}$$

and therefore the average failure rate (AFR) of survivors from time *0* to *t* is:

$$AFR(t) = \frac{1}{t}\int_0^t h(t')dt' = \frac{-\ln(1 - F(t))}{t} \tag{A.6.3}$$

The normal unit of AFR is the inverse of time, such as h^{-1}. 1 FIT is defined as 10^{-9} AFR. Probability plotting (Weibull recommended) of time-to-breakdown t_{bd} gives distribution parameters needed to calculate the AFR. It is recommended that Weibull statistics be used in the case of ultra-thin gate dielectrics since it has been shown experimentally that it correctly predicts area scaling and the thickness dependence on Weibull slope. The distribution parameters are the scale parameter η and shape parameter β for the Weibull distribution. Table 4 shows the equations for AFR calculation for the Weibull distribution.

Once the distribution parameters are known from the probability plotting, AFR can be calculated based on the equations in Table 4. An easy way to do this is to use software that provides the built-in PDF *f(t)*, CDF *F(t)* functions for both Lognormal and Weibull distributions.

Area scaling is another factor for AFR calculation. Assuming the defect distribution is random across the area (could also be randomly distributed across the field oxide edge), the reliability function *R(t)* is:

$$R(t) = 1\text{-}F(t) = e^{-D_A A} \tag{A.6.4}$$

where *A* is the device area and D_A is the defect density.

For area A_1 and A_2, we have:

$$(1\text{-}F(t)_{A1})^{1/A1} = (1\text{-}F(t)_{A2})^{1/A2} \tag{A.6.5}$$

and therefore,

$$AFR(t_1, t_2) = \frac{-\ln(1 - F(t)_{A_1})}{t} = \frac{-\ln(1 - F(t)_{A_2})}{t}\frac{A_1}{A_2} = AFR_{A_2}\frac{A_1}{A_2} \tag{A.6.6}$$

A similar analysis can be performed for the defect density associated with edge or perimeter defects:

$$R(t) = 1\text{-}F(t) = e^{-D_l L} \tag{A.6.7}$$

where *L* is the length of the device perimeter and D_l is defect density. More details can be found in [1, 2].

A.6 Determining Failure Rates – FIT Calculation (cont'd)

Table 4 — Reliability Characteristics for Weibull Distributions

Distribution Parameters	Weibull
PDF, *f(t)*	$\frac{\beta t^{\beta-1}}{\eta^{\beta}} exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right]$
CDF, Unreliability, *F(t)*	$1 - exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right]$
Instantaneous failure rate, *h(t)*	$\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}$
AFR	$\frac{1}{t}\left(\frac{t}{\eta}\right)^{\beta}$

References for A.6:

1. M. Modarres *et al.*, *Reliability Engineering and Risk Analysis*, CRC PRESS 2009.
2. W. Hunter, "The Analysis of Oxide Reliability Data," *IEEE International Integrated Reliability Workshop Final Report*, 1998.

A.7 Combining TDDB Distributions to Obtain Overall Reliability

In many cases, the overall reliability of an integrated circuit must be calculated. The Reliability of a Weibull failure distribution, R, is:

$$R(t) = exp(-(t/\eta)^{\beta}) \quad \text{(A.7.1)}$$

The combined reliability of multiple independent failure distributions is the product of the individual distributions [1].

The circuit may have different types of transistors, NMOS, PMOS, decoupling capacitors, and I/O transistors with thicker dielectrics than the core logic. In addition, if transistors in the circuit spend some time experiencing gate-bias stress and some time experiencing drain-bias stress, the overall reliability based on the two different types of stress experienced must be calculated.

As a simplified example, in order to account for gate-bias stress and drain-bias stress of a circuit, assuming 50% of the time is spent in gate-bias and 50% of the time is spent in drain-bias, Table 5 can be used, where indices 1 and 2 can represent two different gate-lengths, two different threshold voltages or core and I/O transistors [2]. For the purpose of this example, gate bias stress and drain-bias stress are considered independent mechanisms. Further study is required to understand the relationship between the two types of stress applied serially or at the same time to a given transistor and is beyond the scope of this standard. Although this table only considers e.g. two different gate-lengths, multiple lines to consider all of the different pieces of an integrated circuit can be included.

A.7 Combining TDDB Distributions to Obtain Overall Reliability (cont'd)

Table 5 — Simplified Method to Obtain the Overall Reliability for an Integrated Circuit

Index	Area	Gate-length	Stress time	63% cum. lifetime	Weibull slope	Reliability equation A.7.1
1 drain-bias	A(1)	L(1)	t/2	ηVd(1)	βVd(1)	*R*Vd(1)
1 gate-bias	A(1)	L(1)	t/2	ηVg(1)	βVg(1)	*R*Vg(1)
2 drain-bias	A(2)	L(2)	t/2	ηVd(2)	βVd(2)	*R*Vd(2)
2 gate-bias	A(2)	L(2)	t/2	ηVg(2)	βVg(2)	*R*Vg(2)
Overall Reliability (product of the individual reliabilities)						Π(*R*)

This methodology can also be used to characterize the overall low percentile reliability for cases described in A.8 where η_F and β_F are obtained from the low percentile portion of a non-Weibull failure distribution.

References for A.7:

1. R.S. Burlington *et al., Handbook of Probability and Statistics with Tables,* 1970, pp. 336-7.
2. B.E. Weir *et al.,* "Utilizing a Thorough Understanding of Critical Aging and Failure Mechanisms in finFET Technologies to Enable Reliable High Performance Circuits," *Proc. International Reliability Physics Symposium,* 2019, 5B.4, pp. 1-5.

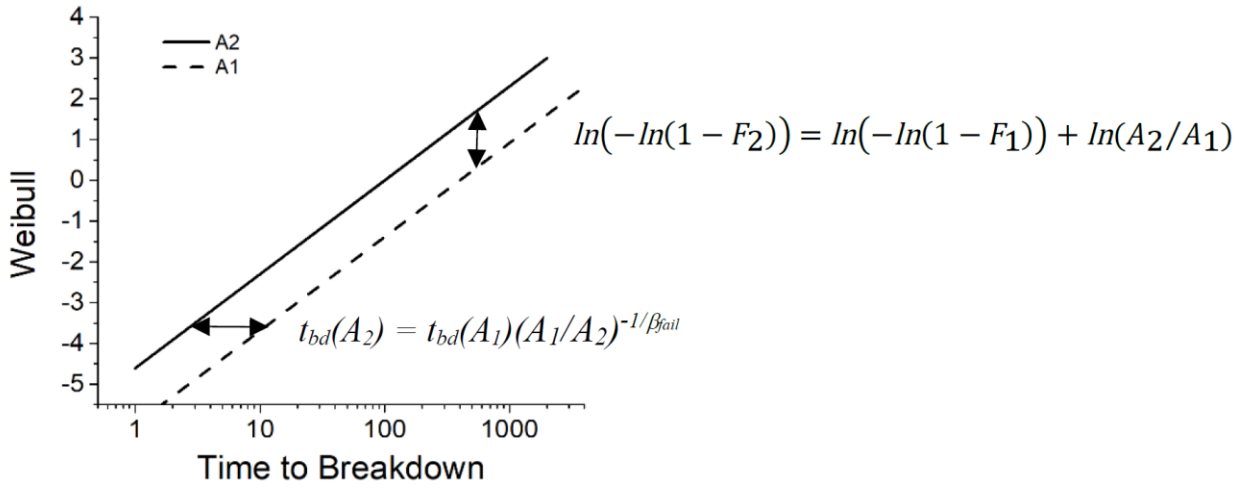
A.8 Non-Weibull Characteristics for TDDB Distributions due to Progressive Breakdown

In addition to the example shown in A.5 of a bimodal distribution with extrinsic and intrinsic components, other common non-Weibull characteristics can be observed. By using a small-area test structure and sensitive breakdown criteria, it is possible to detect the point where a percolation path through the dielectric is first formed, called t_{1bd}. In addition, if all samples included have minimal thickness variation, the distribution is a Weibull distribution with a single slope. However, in actual practice, many dielectric breakdown distributions with breakdown criteria specified at higher levels of current are a combination of cases where the formation of an initial percolation path is detected along with cases where the breakdown is not detected until the initially formed percolation path has widened to allow higher current flow – called progressive breakdown. In that case, the time-to-fail (t_{fail}) is equal to the sum of the two stages; time to first breakdown, t_{1bd}, and the additional time for the percolation path to widen to allow higher current flow t_{pbd} ($t_{fail} = t_{1bd} + t_{pbd}$).

A further complication is that in some cases, additional percolation paths form, and the time to reach the failure current, t_{fail} is a result of the formation of multiple percolation paths. When a t_{fail} distribution is composed of some structures which reached the breakdown criterion at t_{1bd}, and others which reached the breakdown criterion after progressive breakdown or after multiple percolation paths were formed, the resulting distribution is not a single-sloped Weibull distribution but a curve, as shown in Figure 10 [1, 2].

A.8 Non-Weibull Characteristics for TDDB Distributions due to Progressive Breakdown (cont'd)

An important factor to keep in mind for this type of non-Weibull failure distribution is as follows: For a single-sloped Weibull distribution, test results from two different gate-area structures will match when shifted either vertically or horizontally, as shown in Figure 8:



NOTE When TDDB data can be described by a single-sloped Weibull distribution, test results from two gate areas can be matched by shifting horizontally using the equation ($t_{bd}(A_2) = t_{bd}(A_1)(A_1/A_2)^{-1/\beta_{fail}}$) or vertically using the equation ($ln(-ln(1-F_2)) = ln(-ln(1-F_1)) + ln(A_2/A_1)$).

Figure 8 — Weibull Failure Distributions for Two Areas

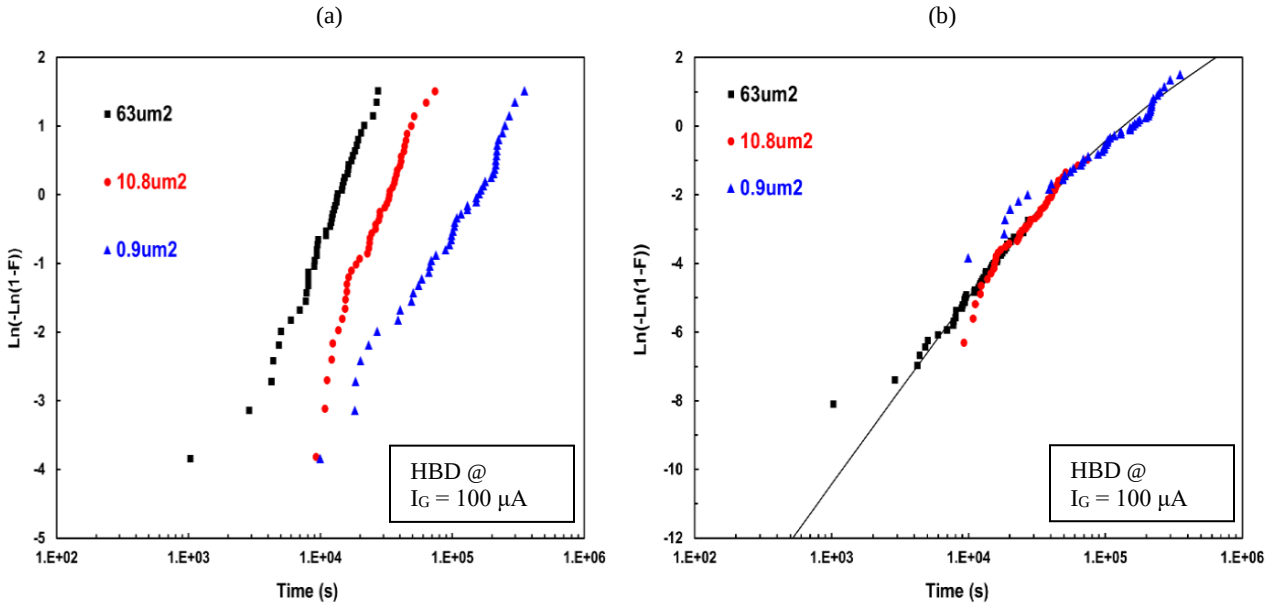
However, when the test data does not follow a single-sloped Weibull distribution, it can only be matched by shifting vertically. As long as the weakest-link property applies and there is a uniform distribution of failure sites in the dielectric area, equation A.8.1 can be used for area-scaling.

$$ln(-ln(1-F_2)) = ln(-ln(1-F_1)) + ln(A_2/A_1) \qquad \text{(A.8.1, also 4.2)}$$

F_1 and F_2 are failure distributions from two sets of test structures with different areas, and A_1 and A_2 are the associated test structure dielectric areas.

A.8 Non-Weibull Characteristics for TDDB Distributions due to Progressive Breakdown (cont'd)

An example of area-dependent data shifted vertically is shown in Figure 9.



NOTE Example of results from a set of tests experiencing progressive breakdown which does not exhibit a single Weibull slope and can be fit by Equation A.8.2. Data from larger-area testers shown in (a) were shifted down in (b) based on Equation A.8.1; $(ln(-ln(1-F_2)) = ln(-ln(1-F_1)) + ln(A_2/A_1))$.

Figure 9 — Multiple Area Test Results Exhibiting Progressive Breakdown

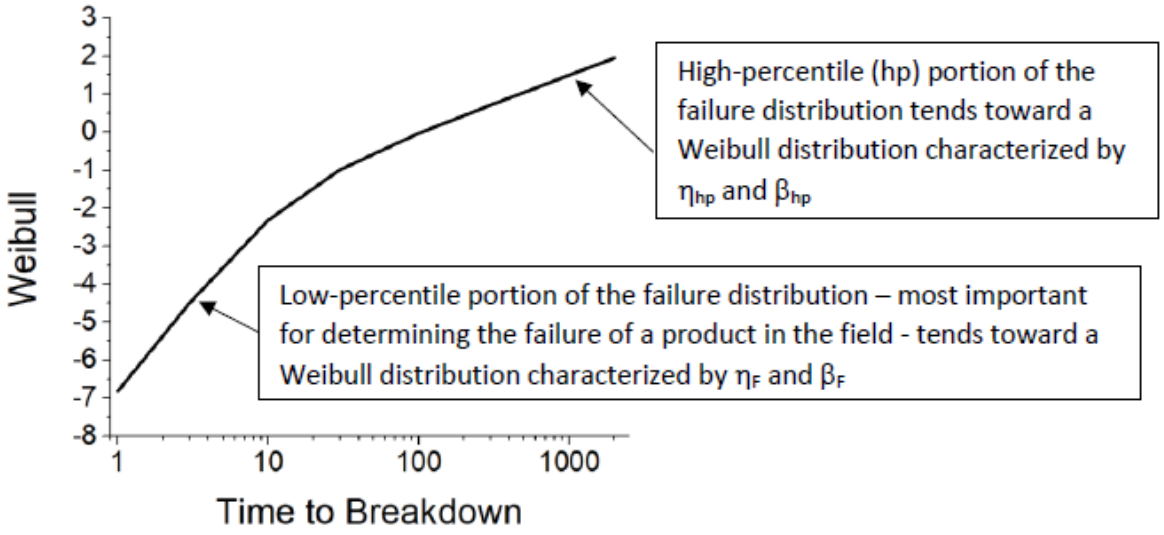
Once the combined failure distribution from multiple areas is obtained, it is commonly observed that the resulting t_{fail} distribution can be described as follows: at higher percentiles, the distribution converges to a shallower slope, whereas at lower percentiles, (most relevant to product reliability criteria), it converges to a steeper slope. These t_{fail} distributions have been studied extensively using Monte Carlo simulations, and a simplified way to analyze such data with only four fitting terms has been developed [3].

$$F_{fail} = 1 - exp\left\{-\left[\frac{\left(\frac{t}{\eta_F}\right)^{\beta_F}\left(\frac{t}{\eta_{hp}}\right)^{\beta_{hp}}}{\left(\frac{t}{\eta_F}\right)^{\beta_F}+\left(\frac{t}{\eta_{hp}}\right)^{\beta_{hp}}}\right]\right\} \quad \text{(A.8.2)}$$

This equation captures the distributions of experimental time-to-fail at a specified failure current level as discussed above. The modal value, η_{hp} and the Weibull slope, β_{hp}, determine the high-percentile section of the combined distribution. The low-percentile section of the combined distribution is dominated by the modal value, η_F, and Weibull slope, β_F.

A.8 Non-Weibull Characteristics for TDDB Distributions due to Progressive Breakdown (cont'd)

Before attempting to fit data using the four-parameter fitting method, observation of progressive breakdown should be verified. Progressive breakdown can be verified by observing the current vs. time characteristic. If the failure condition is determined by I_{fail} as defined in 4.3.2.d and for some samples the current simply increases beyond I_{fail} without a sudden increase in current, the distribution is a progressive breakdown distribution. If all samples show an increase of >5× in I_{meas} at V_{stress} or V_{sense} at the time of breakdown, the distribution is not considered a progressive breakdown distribution and the fitting method in this annex should not be used.



EXAMPLE The failure distribution is obtained by using an I_{fail} criterion where some devices experience first soft-breakdown, while others experience progressive breakdown or multiple-spot breakdown. This distribution can be fit using equation A.8.2 with four parameters: η_{hp} and β_{hp}, which describe the high-percentile portion of the distribution, and η_F and β_F, which describe the low percentile portion of the distribution.

Figure 10 — Example of a Failure Distribution Fit by Equation A.8.2

A.8 Non-Weibull Characteristics for TDDB Distributions due to Progressive Breakdown (cont'd)

In order to determine these four parameters, find η_{hp} and β_{hp} from the higher percentile values of the data plot as a first estimate. Next, use a fitting algorithm to determine η_F and β_F from the entire data plot. Adjustments can be made to η_{hp} and β_{hp} in order to achieve an overall good fit to the data.

Maximum Likelihood Estimation (MLE) is recommended to determine the four parameters of equation A.8.2 for a given dataset [4-8]. These four parameters can only be determined accurately for a large dataset (e.g. >1000 datapoints). For example, a dataset consisting of 2115 datapoints was fit with MLE [9] and β_F was determined. When ~200 datapoints were chosen randomly from the 2115 datapoints and the smaller distributions were fit, β_F ranged from 0.7× to 1.7× the value obtained from the larger dataset. This range of β_F could result in a time to failure at spec. conditions of 0.017× to 59× the time obtained from the larger dataset, which indicates that 200 datapoints is not sufficiently large. In order to justify a β_F value >1, a dataset at a single voltage and temperature with more than 1000 datapoints is recommended. For measurements at additional voltages and temperatures used to determine voltage and temperature acceleration, smaller datasets can be used.

In order to obtain a time-to-fail for a given failure fraction of a product in the range of 1000 ppm to 1 ppm, the voltage projection procedure described in 4.8 from high stress voltages to low use voltages can be performed using η_F and β_F from the low-percentile portion of the combined failure distribution. If needed, temperature acceleration as described in 4.3.1 and 4.8 can also be performed using η_F and β_F from the low-percentile portion of the combined failure distribution.

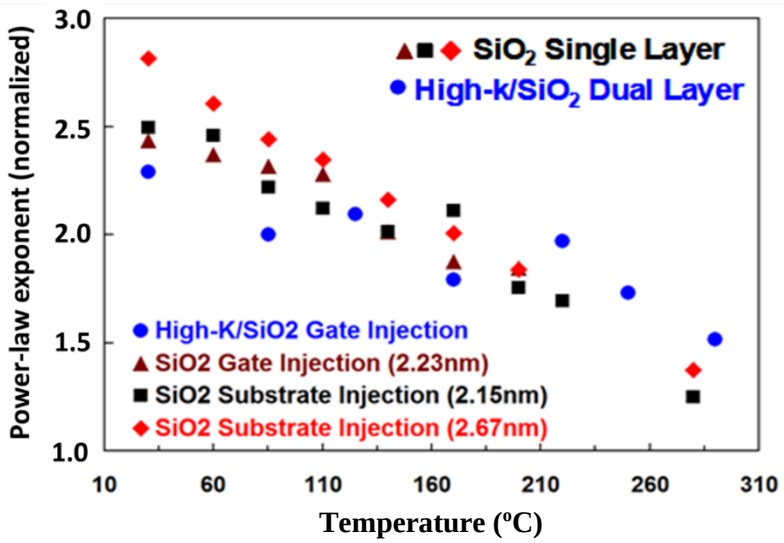
References for A.8:

1. Ernest Wu *et al.,* "Statistical and Voltage Scaling Properties of Post-Breakdown for Ultra-Thin-Oxide Pfets in Inversion Mode," *Proc. International Reliability Physics Symposium*, 2006, pp. 54-62.
2. Kaustubh Joshi *et al*., "A Statistical Learning Model for Accurate Prediction of Time-Dependent Dielectric Degradation for Low Failure Rates," *Proc. International Reliability Physics Symposium*, 4E.4, 2019.
3. Santi Tous *et al.,* "A Compact Model for Oxide Breakdown Failure Distribution in Ultrathin Oxides Showing Progressive Breakdown," *IEEE Electron Device Letters*, 2008, Vol. 29, No. 8, pp. 949-951.
4. In Jae Myung "Tutorial on maximum likelihood estimation," *Journal of Mathematical Psychology*, 2003, Vol. 47, pp. 90-100
5. Baozhen Li *et al.,* "Application of Three-Parameter Lognormal Distribution in EM Data Analysis," *Microelectronics Reliability*, 2006, Vol. 46, pp. 2049-2055
6. W.Q. Meeker and L.A. Escobar, *Statistical Methods for Reliability Data*, Wiley; 1998.
7. J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, Wiley; 1982 2nd Edition pp. 522-532.
8. Jay L. Devore, *Probability and Statistics for Engineering and the Sciences*, Duxbury; 1999 5th Edition pp. 267-273.
9. Maximum Likelihood Estimation of equation A.8.2 was performed using a Python script available on GitHub (GitHub - robinsdp/weibull_four_parameter: Calculates the Maximum Likelihood Regression Coefficients for a Custom 4 Parameter Weibull Distribution) by Davis Robinson, Data Scientist with Vantage Risk.

A.9 Interdependence of Temperature and Voltage Acceleration Factors

Both acceleration parameters, n in the power-law model and γ_V or γ_E in the exponential-models have a temperature dependence [1-3]. In practice, by performing the voltage-dependent measurements at the use temperature, this complication can be avoided.

While a simplified model, commonly used in the industry, assumes that $t_{bd\_63\%} \propto V^{-n}$, the correct equation is: $t_{bd\_63\%} \propto V^{-n(T)}$. If predicting the voltage and temperature dependence over a wide temperature range, the voltage acceleration will need to be re-evaluated in each specific temperature range, as shown in Figure 11 [a].



NOTE $t_{bd\_63\%}$ power-law exponents (n) vs. temperature for high-κ/SiO_2 bilayers and SiO_2 dielectrics demonstrating the need to evaluate n in the specific temperature range of interest.

[a] This figure (Figure 11) was reproduced from Figure 2 in Ernest Y. Wu and Jordi Suñé, "Generalized Hydrogen Release-Reaction Model for the Breakdown of Modern Gate Dielectrics," *J. Appl. Phys.* 114, 014013, 2013, with permission of AIP Publishing.

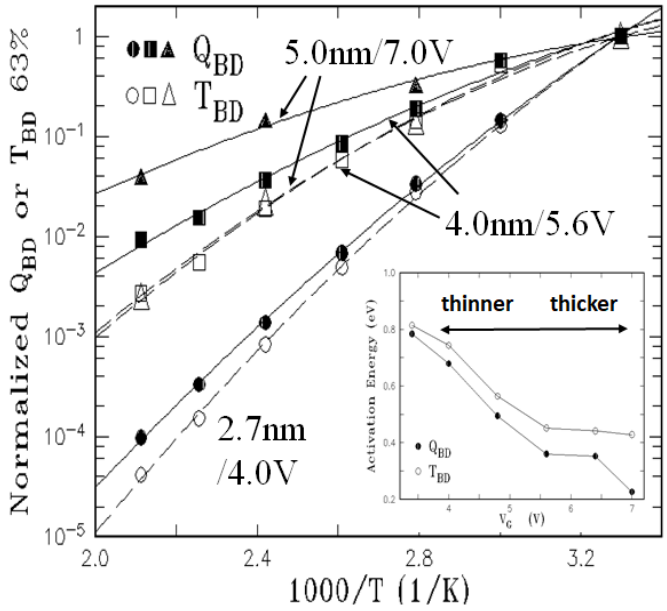
Figure 11 — Power-law Exponents vs. Temperature

In addition, over a range of voltages, the temperature dependence does not remain constant [5-7]. This behavior becomes more pronounced as gate dielectric thickness decreases. Some research has found that the following equation can be used to more accurately fit TDDB data over a range of temperatures [4].

$$t_{bd} \propto \exp\left(\frac{a(V)}{T} + \frac{b(V)}{T^2}\right) \quad \text{(A.9.1)}$$

The activation energy increases as the voltage decreases, as shown in Figure 12. The degree of change observed is technology and process specific. Since CVS testing is usually performed at voltages higher than use voltage, determining the activation energy at a test voltage which gives $t_{bd\_63\%}$ ~10 s to 1000 s will result in a conservative value for the activation energy.

A.9 Interdependence of Temperature and Voltage Acceleration Factors (cont'd)



NOTE The figure shows the non-Arrhenius behavior of TDDB over a range of temperatures. The inset shows how the activation energy changes as temperature-dependence is measured over different voltage ranges for different gate dielectric thicknesses.

[a] The rights to use this figure (Figure 12) in this document were purchased by JEDEC from Elsevier Publishing (Figure 6 in [4]).

Figure 12 — Non-Arrhenius Behavior of TDDB

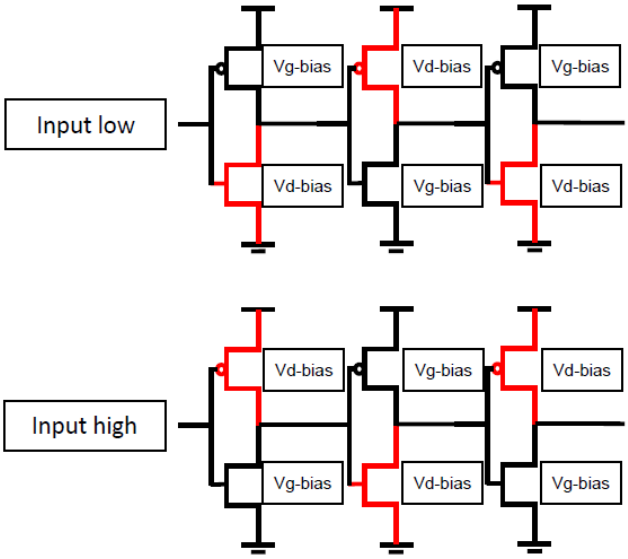
References for A.9:

1. E. Wu *et al.,* "Temperature Dependence of TDDB Voltage Acceleration in High-κ/ SiO_2 Bilayers and SiO_2 Gate Dielectrics," *International Electron Devices Meeting*, 2012, pp. 28.5.1-28.5.4.
2. Ernest Y. Wu and Jordi Suñé, "Generalized Hydrogen Release-Reaction Model for the Breakdown of Modern Gate Dielectrics," *J. Appl. Phys.* 114, 014013, 2013.
3. Ernest Y. Wu, "Facts and Myths of Dielectric Breakdown Processes – Part 1: Statistics, Experimental, and Physical Acceleration Models," *IEEE Trans. Electron Dev.*, Vol. 66, No. 11, pp. 4523-4534, 2019
4. E. Wu *et al.,* "Interplay of Voltage and Temperature Acceleration of Oxide Breakdown for Ultra-Thin Gate Oxides," *Solid State Electronics,* Vol. 46 pp. 1787-1798, 2002
5. J. McPherson, "Stress Dependent Activation Energy," *IEEE International Reliability Physics Symposium Proceedings,* 1986, pp. 12-18.
6. J. Suehle *et al.,* "Field and Temperature Acceleration of Time-Dependent Dielectric Breakdown in Intrinsic Thin SiO," *IEEE International Reliability Physics Symposium Proceedings*, 1994, pp. 120-125.
7. R.-P. Vollertsen and W.W. Abadeer, "Upper Voltage and Temperature Limitations of Stress Conditions for Relevant Dielectric Breakdown Projections," *Quality and Reliability Engineering International*, Vol. 11, pp. 233-238, 1995.

A.10 TDDB Results After AC Stress

A.10.1 Single Transistor AC Measurements

This section discusses unipolar AC stress with on/off stress of a single terminal – either Vg or Vd. This captures the two primary stress conditions experienced by an inverter as shown in Figure 13. Two other stress conditions are beyond the scope of this standard. The first is bipolar AC stress, e.g. where gate-bias stress transitions back and forth between inversion (on-state during normal operation) and accumulation (not part of normal operating conditions). The second is inversion AC stress where both Vg and Vd are applied simultaneously (e.g. during a transition from input low to high or from high to low).



NOTE Delay chain showing that half of the transistors experience Vd-bias (shown in red) while the other half experience Vg-bias. When the input alternates between high and low, each transistor alternates between Vg-bias and Vd-bias.

Figure 13 — Effect of Input on Transistor Bias in a Delay Chain

The range of operating conditions of each circuit application will need to be understood in order to determine whether or not a calculation of failure fraction at a given time under AC stress is valid. For a given design, toggle activity and approximate frequency should be known or estimated in order to determine an approximate fraction of transistors under AC or DC stress. For example, digital circuits may have Value Change Dump (VCD), Fast Signal Database (FSDB) or Switching Activity Interchange Files (SAIF) which contain the exact toggle conditions for a given functional mode. Alternatively, the toggle activity may be estimated by the design team or provided by the customer. This toggle activity information is routinely used for IR-drop and power estimates and may be used to estimate the fraction of transistors at frequencies over 1 MHz vs. those under semi-static conditions.

Transistor-level studies where an AC waveform is applied to the gate or drain of a transistor are the primary source of published information about AC TDDB. These studies have shown that AC stress above 1MHz results in a longer lifetime than DC CVS [1-3]. Ref. 1 shows how to test an individual or array of transistors at high frequency by including an on-die ring oscillator to provide the AC waveform. The frequency at which the AC lifetime exceeds the DC lifetime depends on the gate dielectric fabrication process and the stress voltage [1-6]. In addition, the voltage-dependence, area-dependence, temperature-dependence and statistical-dependence may change with AC stress, so that each of these will need to be determined in order to determine the failure fraction at a given time under AC stress.

A.10.1 Single Transistor AC Measurements (cont'd)

Research work is inconclusive at this time regarding voltage dependence, area dependence, statistical dependence and temperature dependence for AC vs. DC stress. In addition, there is insufficient data at this point to enable specific projections over a range of duty cycles, so a 50% duty cycle is recommended for AC measurements and projections. If a study of AC/DC ratio as a function of duty cycle is performed in the future, projection of AC lifetime over a range of duty cycles may be possible.

When comparing AC and DC measurements at the same temperature, voltage and gate dielectric area, it is important to ensure that both measurements are at the same junction temperature. For unipolar AC drain-bias TDDB stress, the junction temperature may be lower than under DC drain-bias stress.

In addition, a display or plot of the actual waveform (e.g. square-wave or sinusoidal wave, etc.) measured at the DUT under AC stress should be provided for the assessment of applicability of AC stress data to a circuit operation condition.

A.10.2 Cautions about Measurement of Gate Dielectric Breakdown in Ring Oscillator Elements:

While measurements performed on transistors or arrays of transistors using an AC input (using an external or on-die ring-oscillator stimulus) allows clear separation of NMOS and PMOS results, measurements performed on ring oscillator elements reach higher frequencies and may more easily test larger areas. Long chains and low capacitive load should be used to minimize the stress time spent in the transition region where both gate and drain bias are applied simultaneously, and where hot-carrier degradation can be significant. Aging due to bias-temperature instability (BTI) always occurs under gate-bias TDDB stress conditions. When stressing a ring oscillator, aging will result in a waveform which becomes gradually more and more rounded in shape, and therefore, the stress will gradually become less severe over time.

In addition, since both NMOS and PMOS within the ring oscillator are stressed, ring oscillator AC measurements are only useful for a given technology if either NMOS or PMOS has a shorter lifetime over the range of voltage, temperature and area used for the set of AC measurements. Drain-bias TDDB tends to result in much harder breakdown than gate-bias TDDB, so that observation of significant differences in breakdown severity can signal, for example, a mix of NMOS gate-bias TDDB and PMOS drain-bias TDDB within a set of ring oscillator TDDB measurements. It is likely that less severe breakdowns will not be detected at all either due to background leakage in long-chain ROs and only more severe breakdowns will be detected if using a ring oscillator test structure. For these reasons, for many technologies, only transistor measurements with AC input stress can be used to characterize gate dielectric breakdown under AC stress.

References for A.10:

1. M. Arabi *et al.*, "New Insights on Device Level TDDB at GHz Speed in Advanced CMOS Nodes" IEEE Transactions on Device and Materials Reliability, 2019 Vol. 19, Issue 2
2. Hyunjin Kim *et al.,* "A Systematic Study of Gate Dielectric TDDB in FinFET Technology" *Proc. International Reliability Physics Symposium*, 2018, 4A.4 pp. 1-4.
3. C.H. Yang *et al.,* "The Physical Explanation of TDDB Power Law Lifetime Model Through Oxygen Vacancy Trap Investigations in HKMG NMOS FinFET Devices," *Proc. International Reliability Physics Symposium*, 2017, 3C.4 pp. 1-6.

References for A.10 (cont'd)

4. A. Kerber *et al.*, "Impact of Charge Trapping on the Voltage Acceleration of TDDB in Metal Gate/High-k N-channel MOSFETs," *Proc. International Reliability Physics Symposium*, 2010, pp. 369-372.
5. P. J. Liao *et al.*, "A New On-state Drain-bias TDDB Lifetime Model and HCI Effect on Drain-bias TDDB of Ultra Thin Oxide," *Proc. International Reliability Physics Symposium*, 2008, pp. 210-214.
6. P. S. Chen *et al.*, "AC TDDB Analysis for HK/IL Gate Stack Breakdown and Frequency-dependent Oxygen Vacancy Trap Generation in Advanced nodes FinFET Devices by SILC Spectrum Methodology," *Proc. International Reliability Physics Symposium*, 2022, 11A.4 pp. 1-6.
7. Xinwei Yu *et al.*, "GHz AC to DC TDDB Modeling with Defect Accumulation Efficiency Model" *Proc. International Reliability Physics Symposium*, 2023, 4C.3 pp. 1-6.

A.11 Evaluating TDDB for the Range of Dielectric Thicknesses Which Occur in a Process

It has been demonstrated that dielectric thickness (th_{diel}) variation can broaden the measured t_{bd} distributions with an apparent shallower Weibull slope, thus leading to an unrealistic reliability projection [1,2]. In reliability qualification, this effect may be taken into account by evaluating the TDDB lifetime not only for nominal thickness but also for the minimum thickness. At least one lot with the minimum thickness along with other lots of nominal thickness values are recommended for TDDB evaluation. The minimum thickness value can be set at $th_{diel} = \mu - 3\sigma$ where μ is the mean value and σ is the standard deviation.

The lifetime projection can be made using the two different methodologies: the combination-based methodology [2,3] and the time-dependent clustering model [4]. In the time-dependent clustering model the cumulative distribution function (CDF) is given as:

$$F_{tot}(t) = 1 - \left[1 + \frac{1}{\alpha}\left(\frac{t}{\eta}\right)^{\beta}\right]^{-\alpha} \qquad \text{(A.11.1)}$$

The parameter, α, describes the degree of non-uniformity. As $\alpha \to \infty$, equation A.11.1 recovers the analytic form of the Weibull distribution, where β and η are the characteristic slope and time to breakdown at 63.2%. The two parameters, β and $\eta(t_{bd\_63\%})$, in equation A.11.1 are the same as those defined in the Weibull distribution so that equation 4.1 is still applicable. In general, at higher percentiles, equation A.11.1 deviates from a Weibull distribution whereas at low percentiles it approaches a straight line in a Weibit plot. This model is developed to deal with the situation where the failure distribution becomes non-Weibull or convexly curved at higher percentiles due to thickness variation effects [1-4]. As a result, the measured failure distributions do not merge together according to the Poisson area-scaling relation of equation A.8.1.

The area-transformation of any two failure distributions can be done as follows [4]:

$$F_2 = 1 - \left[1 + \frac{A_2}{A_1}\left((1 - F_1)^{-\frac{1}{\alpha}} - 1\right)\right]^{-\alpha} \qquad \text{(A.11.2)}$$

Similar to equation A.11.1, as $\alpha \to \infty$, equation A.11.2 approaches the conventional Poisson area-scaling of equation A.8.1. In practice, the experimental CDF measured from any two areas can be translated to a reference area of relatively small area by obtaining an optimal α value to merge the translated distributions into a universal straight-line. In this case, the Poisson area-scaling relationship is restored at low percentiles. Then, the η and β values can be obtained by fitting the merged distributions [4].

A.11 Evaluating TDDB for the Range of Dielectric Thicknesses Which Occur in a Process (cont'd)

Alternatively, a combination-based methodology [2,3] can be used to obtain the failure CDF in equation A.11.3 as follows:

$$F_{tot}(t) = \int p(th_{diel}; \mu, \sigma) F_{WB}(t, th_{diel}) dth_{diel} \quad \text{(A.11.3)}$$

where t is the time-to-breakdown and th_{diel} is the dielectric thickness. $p(th_{diel};\mu,\sigma)$ is the probability density function (PDF) of a th_{diel} distribution with μ and σ being the mean and the standard-deviation. $F_{WB}(t)$ is the Weibull CDF for time-to-breakdown in equation A.11.4.

$$F_{WB}(t) = 1 - \exp\left(-\left(\frac{t}{\eta(th_{diel})}\right)^{\beta(th_{diel})}\right) \quad \text{(A.11.4)}$$

and the th_{diel} distribution in the form of a normal distribution:

$$p_{Norm} = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(th_{diel}-\mu)^2}{2\sigma^2}\right) \quad \text{(A.11.5)}$$

The thickness dependence of β (Weibull slope) and η (63% cum. time to failure) in equation A.11.4 are given in equation A.11.6 and equation A.11.7. The fitting parameters, a, b, c, and D can be obtained from the measured t_{bd} distributions with at least two thickness values with minimized thickness variation.

$$\beta(th_{diel}) = a + b * th_{diel} \quad \text{(A.11.6)}$$

$$\eta(th_{diel}) = D * exp(c * th_{diel}) \quad \text{(A.11.7)}$$

The integral in equation (A.11.3) can be solved by using a Monte Carlo approach [2] or a numerical approach [3]. The combination-based approach described above requires six parameters [2] which need to be determined independently from separate experiments, but this methodology is generally valid.

On the other hand, the time-dependent clustering model only contains three parameters so that it is easier to use. For FEOL gate dielectrics with relatively small thickness variation across a wafer, the time-dependent clustering model is usually applicable and a valid approach. For FEOL gate dielectric TDDB, the equivalence of these two methodologies have been demonstrated using identical FEOL SiO_2 TDDB datasets [5].

However, in cases with excessive variability issues, it may not be possible to use the time-dependent clustering model to merge the experimental CDF data together to restore the Poisson area-scaling relationship. Then, a generalized combination-based methodology can be used [6].

References for A.11:

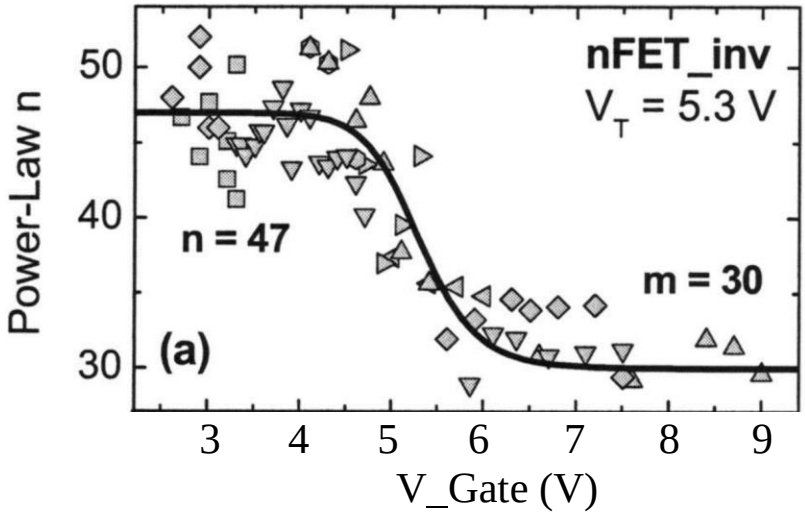
1. B.E. Weir *et al.,* "Gate Oxides in 50 nm Devices: Thickness Uniformity Improves Projected Reliability," *International Electron Devices Meeting*, p. 437, 1999.
2. E.Y. Wu *et al.,* "Weibull Breakdown Characteristics and Oxide Thickness Uniformity," *IEEE Trans. on Electron Devices,* vol. 47, pp. 2301-2309, 2000.
3. T. Pompl *et al.,* "Modeling of Substrate Related Extrinsic Oxide Failure Distributions," *Proc. International Reliability Physics Symposium*, pp. 393-403, 2002.
4. Ernest Y. Wu *et al.,* "Modeling of Time-dependent Non-uniform Dielectric Breakdown Using a Clustering Statistical Approach," *Applied Physics Letters,* vol. 103, 152907, 2013.

References for A.11 (cont'd)

5. E.Y. Wu *et al.,* "TIME-dependent Clustering Model Versus Combination-Based Approach for BEOL/MOL and FEOL Non-uniform Dielectric Breakdown: Similarities and Disparities," *Proc. International Reliability Physics Symposium*, pp. 5B.2.1-7, 2014.
6. Ernest Wu *et al.,* "A Flexible and Inherently Self-Consistent Methodology for MOL/BEOL/MIMCAP TDDB Applications with Excessive Variability-Induced Degradation," *Proc. International Reliability Physics Symposium* , pp. 2A.2.1-9, 2022.

A.12 Methodology for Voltage Acceleration Factor Projection from the Fowler-Nordheim Regime to the Direct-Tunneling Regime

For FEOL gate dielectric TDDB reliability, in a limited range of dielectric thicknesses, one may deal with a situation in which stress voltages are large enough so that conduction mechanisms are dominated by the Fowler-Nordheim (F-N) tunneling mechanism whereas the use conditions are specified at lower voltages corresponding to the direct tunneling (DT) regime (tunneling mechanisms are described in Annex C). It is reported that the power-law exponents obtained in the F-N regime are smaller than those of the DT regime [1].



[a] The rights to use this figure (Figure 14) in this document were purchased by JEDEC from IEEE Publishing (Figure 2(a) in [1]).

Figure 14 — Voltage Acceleration Factor vs. Stress Voltage

A.12 Methodology for Voltage Acceleration Factor Projection from the Fowler-Nordheim Regime to the Direct-Tunneling Regime (cont'd)

Therefore, the projection using the power-law exponents in the F-N regime to the use-voltage in the DT regime would result in a conservative result. To provide a more realistic life-time projection, a combined t_{bd} power-law model is proposed as follows [1].

$$t_{bd} \sim (\frac{V}{V_T})^{-n} + (\frac{V}{V_T})^{-m} \quad \text{(A.12.1)}$$

where *n* and *m* are the power exponents derived from the stress data in the DT and F-N regimes, respectively, with a voltage fitting parameter, V_T (e.g. V_T = 5.3V).

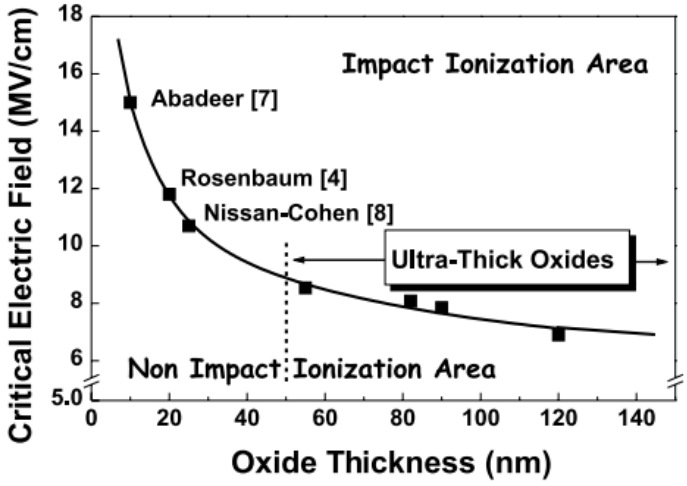
As expected, in the DT regime, the projected lifetime using this combined t_{bd} power-law model coincides with the projection results of using only the first term in equation A.12.1.

References for A.12:

1. R. Duschl and R.-P. Vollertsen, "Voltage Acceleration of Oxide Breakdown in the Sub-10 nm Fowler-Nordheim and Direct Tunneling Regime," *IEEE International Integrated Reliability Workshop, Final Report*, pp.44-48, 2005.

A.13 Effects of Impact Ionization on Gate Dielectric Breakdown Testing

As shown in Figure 15, impact ionization (SiO_2 bandgap ionization by carriers with energies exceeding 9 eV) can occur at fields as low as 6.5 MV/cm for ultra-thick oxides [1]. Another report states that band-gap ionization will begin to occur at ≈7.5 MV/cm only for oxide thicknesses in the 30 nm to 40 nm range [2]. When measuring time to breakdown vs. voltage and the voltage range used overlaps the impact-ionization regime, voltage acceleration and Weibull slope can be distorted [3-4]. In particular, the voltage acceleration factor measured in the impact ionization regime at high fields can be much larger than that measured in fields below the impact ionization threshold [5-6].

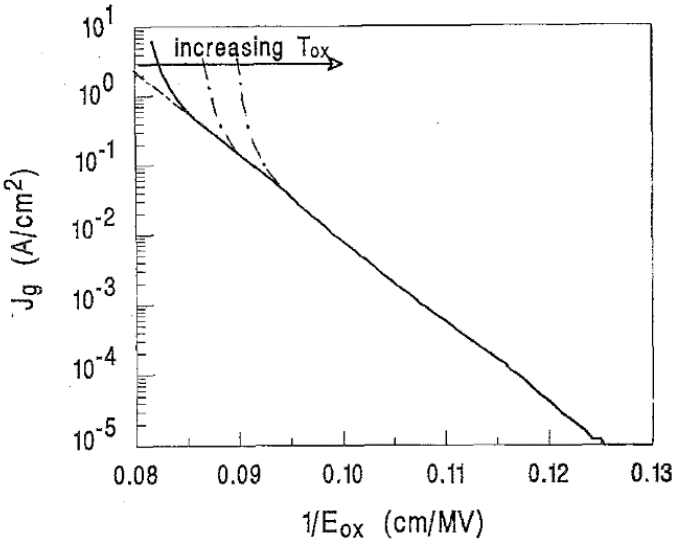


[a] The rights to use this figure (Figure 15) in this document were purchased by JEDEC from Elsevier Publishing (Figure 6 in [1]).

Figure 15 — Impact Ionization Threshold vs. Oxide Thickness

A.13 Effects of Impact Ionization on Gate Dielectric Breakdown Testing (cont'd)

One clear indication of impact ionization is shown in Figure 16. The tunneling current through oxides thicker than 20 nm can be observed to deviate from Fowler-Nordheim tunneling characteristics at electric fields in the 10.5 MV/cm to 11.8 MV/cm range.



NOTE Evidence of the onset of impact ionization in oxides thicker than 20 nm caused by positive charge build-up in the bulk can be observed as a deviation from Fowler-Nordheim tunneling characteristics[6].

[a] The rights to use this figure (Figure 16) in this document were purchased by JEDEC from IEEE Publishing (Figure 7 in [6]).

Figure 16 — Tunneling Characteristics Showing the Onset of Impact Ionization

One solution to avoid the impact ionization regime is to use long-term package-level testing to perform CVS at voltages below the impact-ionization threshold. Alternatively, electric field distortion due to positive charge trapping caused by impact ionization can be calculated and the time to breakdown at a given electric field can be calculated. This technique can be used to determine accurate failure probability distributions or voltage acceleration factors even when the voltage stress range spans the onset of impact ionization [3-4].

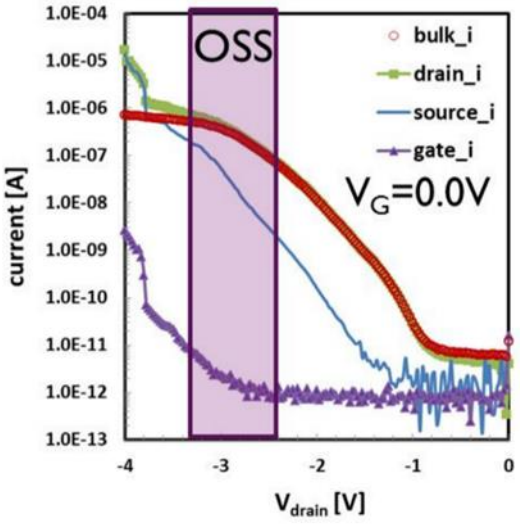
References for A.13:

1. U. Schwalke *et al.*, "Ultra-thick Gate Oxides: Charge Generation and its Impact on Reliability," *Microelectronics Reliability* Vol. 41, 2001, pp. 1007-1010.
2. D. J. DiMaria *et al.*, "Impact Ionization and Positive Charge Formation in Silicon Dioxide Films on Silicon", *Appl. Phys.Lett.* 60 (17), 1992, pp. 2118-2120.
3. A. Aal, "Fast Wafer Level Reliability Assessment of Ultra Thick Oxides Under Impact Ionization Conditions," *2007 IEEE International Integrated Reliability Workshop Final Report*, pp. 117-120.
4. A. Aal, "Comparison Between Gate Oxide Lifetime Models with R_{series} and Trapping Effect Correction in the FN-regime," *2012 IEEE International Integrated Reliability Workshop Final Report*, pp. 99-104.
5. W.W. Abadeer *et al.*, "Correlation Between Theory and Data for Mechanisms Leading to Dielectric Breakdown," *Symposium on VLSI Technology*, 1994, pp. 43-44.
6. E. Rosenbaum *et al.*, "Accelerated Testing of SiO_2 Reliability," *IEEE Transactions on Electron Devices*, vol. 43, 1996, pp. 70-80.

A.14 Drain-bias TDDB Testing and Source-Drain Punch-through

In a transistor, as the drain voltage is increased while gate, source and well are grounded, the source-drain current begins to increase due to drain-induced barrier lowering (DIBL). As the drain-voltage is increased further to a stress voltage, source-drain punch-through, characterized by a steep increase of source current has been observed [1-4].

Even when the stress voltage is kept below the voltage at which punch-through is observed in a drain-bias sweep (see the off-state stress (OSS) regime in Figure 17), and the stress begins with drain current matched to well-current, the source current can increase as the stress proceeds until it matches the drain current. I-V curves performed after such stress (when the stress is stopped before drain-bias TDDB occurs) can show orders of magnitude increases in source-drain current at very low Vg and Vd, indicating that the transistor characteristics have changed dramatically under these stress conditions. An increase in off-current as a function of stress time for drain-bias stress has been observed and, in some cases, characterized as an increase in DIBL [5]. For technologies and transistor types in which such a steep increase in matched source-drain current exists at the drain-bias TDDB stress voltage, performing drain-bias stress in the punch-through regime likely involves mechanisms which do not occur under normal operation and could result in a pessimistic lifetime projection.



NOTE The figure shows the result of a drain-bias sweep with source, gate and well grounded. Punch-through is observed at ~-3.7V and the authors describe the voltage ranging from ~2.5V to ~3.3V as the off-state stress regime. The voltage at which punch-through occurs depends on the transistor type, technology, and gate-length, and some transistors do not show clear punch-through behavior.

[a] The rights to use this figure (Figure 17) in this document were purchased by JEDEC from IEEE Publishing (Figure 2 in [4]).

Figure 17 — Example of Drain Bias Sweep with Other Transistor Connections Grounded

References for A.14:

1. D. Nminibapiel *et al.,* “Method to Evaluate Off-state Breakdown in Scaled Tri-gate Technologies,” *Proc. International Reliability Physics Symposium*, 9B.1, 2022.
2. M. Koyanagi *et al.,* “Hot-electron-induced Punch-through (HEIP) Effect in Sub-micrometer PMOSFET's,” in *IEEE Transactions on Electron Devices*, vol. 34, no. 4, pp. 839-844, April 1987.
3. J. Franco *et al.,* “Hot Electron and Hot Hole Induced Degradation of SiGe p-FinFETs Studied by Degradation Maps in the Entire Bias Space,” *IEEE International Reliability Physics Symposium (IRPS)*, 2018, pp. 5A.1-1-5A.1-7.
4. M. Cho *et al*., "Off-state Stress Degradation Mechanism on Advanced p-MOSFETs," *2015 International Conference on IC Design & Technology (ICICDT)*, 2015, pp. 1-4.
5. C. Gupta *et al.*, "Impact of Hot-Carrier Degradation on Drain-Induced Barrier Lowering in Multi-fin SOI n-Channel FinFETs With Self-Heating," in *IEEE Transactions on Electron Devices*, vol. 67, 2020, pp. 2208-2212.
6. K. Joshi *et al.*, “A Detailed Comparison of Various Off-state Breakdown Methodologies for Scaled Tri-gate Technologies.” *Proc. International Reliability Physics Symposium*, 2023. 9A.2 pp. 1-6.

Annex B (Informative) Supplemental Sampling Plan Statistics

B.1 Overview

This annex is provided for historic purposes and is applicable to thicker oxides where breakdown detection is possible even for large-area capacitor test structures. The methodology outlined does not account for process variation, which can result in differences in gate oxide breakdown characteristics. These methodologies are not possible to implement when only small-area test structures make it possible to detect dielectric breakdown. For these reasons, other types of sampling plans are employed today as described in 4.8. The terminology used in this section is defined within this section and not synchronized with the terms and definitions in 3.

Typical semiconductor production process defect densities are at levels so low as to be extremely difficult to measure with any degree of precision. Consequently, the primary use of wafer level test structures is in qualitative applications, such as verification of process control or identification of defect sources, rather than as an absolute measure of defect density. However, in certain special applications it may be desirable to extract a statistically valid defect density.

This annex presents a three-step procedure that allows a manufacturer to;

1) determine the maximum defect density that must be demonstrated to ensure that no oxide defects exist on a production die (see B.2),
2) establish a sampling plan that allows a desired defect density of D_0 or less to be demonstrated (see B.3), and
3) estimate defect density given a specific test result (see B.4 and B.5). Examples of how this procedure might be implemented are given in B.6.

Care must be taken when using these procedures to ensure that the sampling plans obtained are practical. In some cases, the total oxide area that must be sampled to prove the desired defect density level is prohibitively large, and the manufacturer should revert to statistical process control techniques rather than direct defect density measurement.

The intent of sampling plan procedures is twofold;

1) to allow a manufacturer to show that no defective oxides exist on production circuits, and
2) to allow tracking of defect density when a problem exists so that it may be effectively reduced.

This is based on the premise that the best technique for managing the reliability implications of defect densities is to eliminate them. More sophisticated (and controversial) techniques are available that attempt to analyze defects and predict operating life failures resulting from those defects.

Two classes of defects can be identified;

1) those whose occurrence is distributed over the entire area of thin oxide, and
2) those whose occurrence is restricted, by their nature, to specific thin oxide edges.

An example of the former is a random particle defect, while an example of the latter is a defect induced by white ribbon thinning. Defect densities can be calculated, monitored, and controlled for each of these types of defects independently. The three most common procedures for controlling this are to monitor the bulk oxide defects lumped together, the gate to source/drain edge effects lumped together, and the gate to field edge effects lumped together.

Annex B (Informative) Supplemental Sampling Plan Statistics (cont'd)

The following procedure describes a technique to control bulk oxide defects, with B.5 describing a modification to this procedure that can be used to monitor both types of edge defects.

B.2 Determining an Acceptable Defect Density Level

Given a circuit with a total gate oxide area of A_c, the probability of finding one or more defects on that circuit is determined by the defect density of the process used to build that circuit. The defect density that results in a probability of 0.95 that no defects are present may be thought of as the defect density that assures, with 95% confidence, that no defective oxides are present. This defect density is called D_0, and is given by the equation:

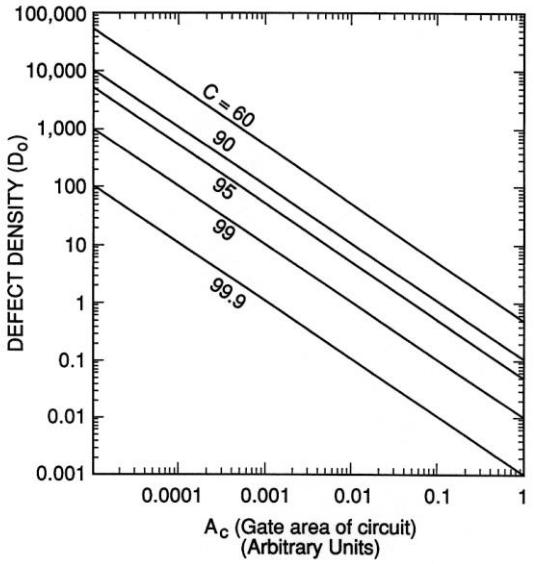
$$D_0 = \frac{-ln\,(0.95)}{A_c} = \frac{0.0513}{A_c} \qquad \text{(B.2.1)}$$

where A_c = Total gate area of circuit (cm^2), and D_0 = Defect density to be proven (defects/cm^2).

Figure 18 shows the D_0 required for circuits with gate areas ranging from 0.00001 to 1 cm^2 to assure defect-free circuits. The lines in the figure show D_0 confidence levels for 60%, 90%, 95%, 99%, and 99.9%.

Equation B.2.1 is based on 0.95 probability of no defects occurring on a circuit, given a Poisson defect distribution. Calculations may be made based on other probability levels using the general equation relating defect density (D_0) to circuit area (A_c) and probability (C) of no failures:

$$D_0 = \frac{-ln\,(C)}{A_c} \qquad \text{(B.2.2)}$$



NOTE D_0 required to assure that a circuit with total gate area A_c is defect free with confidence levels from 60% to 99.9%.

Figure 18 — Required D_0 to Assure Defect-free Circuits

B.3 Sampling Required to Demonstrate Defect Densities

This section describes statistical procedures for determining a sampling plan for demonstrating a given defect density. Consider the following:

D_0 = Acceptable defect density (defects/cm^2)
A_t = Gate area of each test structure (cm^2)
N = Number of test structures sampled

A defect density of D_0 or better can be proven by stressing N test structures, each of gate area A_t, and experiencing zero defect related failures. See A.5 for a description of "defect related failure."

Two procedures are described: one that allows determination of N if A_t is fixed, and one that allows determination of A_t if there are constraints on N.

B.3.1 Finding N Given A_t

Equation B.3.1 gives the relationship between sample size required and test structure area:

$$N \geq \frac{-ln(1-0.95)}{A_t D_0} \quad \text{(B.3.1)}$$

B.3.2 Finding A_t Given N

Equation B.3.2 gives the relationship between test structure area required and sample size

$$A_t \geq \frac{-ln(1-0.95)}{N D_0} \quad \text{(B.3.2)}$$

Equation B.3.1 and equation B.3.2 prove to a 95% confidence level (C) that the actual defect density is less than or equal to D_0, with no defective oxides allowed in a sample of size N. Equation B.3.1 and equation B.3.2 are specific solutions of the general equation:

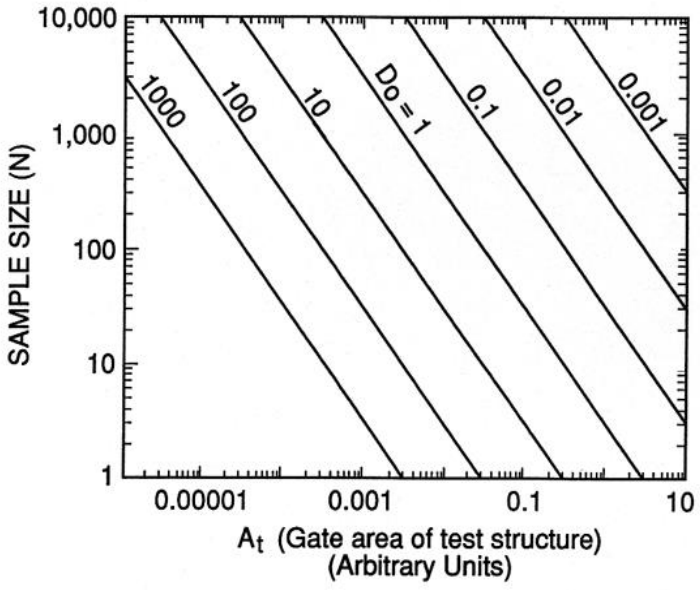
$$C \leq 1 - \sum_{i=Y}^{N} \frac{N!}{i!(N-i)!} [\exp(-D_0 A_t)]^i [1 - \exp(-D_0 A_t)]^{N-i} \quad \text{(B.3.3)}$$

where Y is the minimum acceptable number of non-defective oxides found among the N oxides tested.

NOTE — In equation B.3.1 and equation B.3.2 all oxides must be non-defective, so Y = N.

B.3.2 Finding A_t Given N (cont'd)

Figure 19 illustrates this relationship for seven distinct values of D_0, ranging from 1000 defects/cm^2 to 0.001 defect/cm^2.



NOTE The sample size (N) versus the gate area of the test structure, for D_0 ranging from 1000 cm^{-2} to 0.001 cm^{-2}.

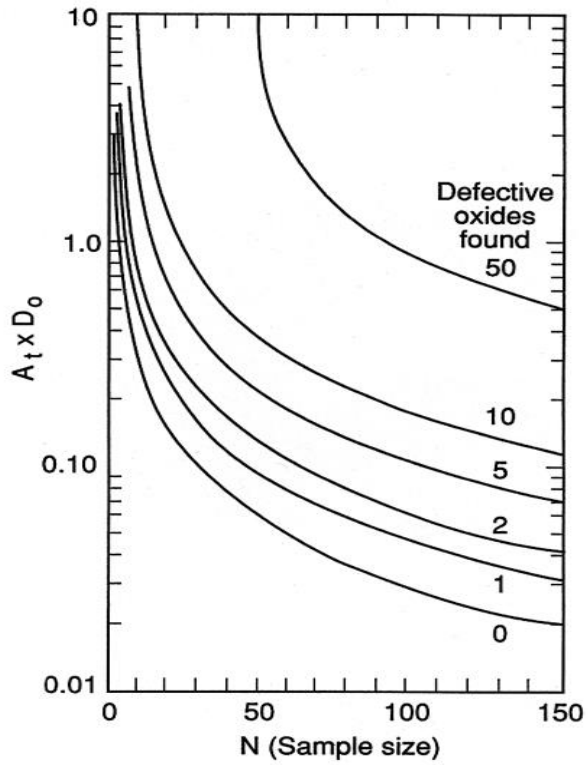
Figure 19 — D_o and Sample Size vs. Gate Area

B.4 Determining Defect Density from a Test Result

B.4 assumes a 95% confidence level that the defect density is D_0 or better and uses equation B.3.3 to generate Figure 20.

Figure 20 may be used to determine a best estimate of defect density given a sample size, N, the number of defective oxides found (curves for 0, 1, 2, 5, 10, and 50 defective oxides are shown), and a test structure with thin oxide area A_t. The vertical axis of the graph is the product of the estimated defect density, D_0, and the test structure thin oxide area, A_t.

To use Figure 20, find the sample size tested on the horizontal axis. Move up to the curve that represents the number of defective oxides found. If there is not a curve for the number found, perform a logarithmic interpolation between the two closest curves. Then, move to the left of the chart and read the value on the vertical axis. Divide this number by the area of the test structure to determine the estimate of defect density.



NOTE Best estimate of the defect density D_0, given a sample size N, the number of defective oxides found, and the area of the thin-oxide test structure A_t.

Figure 20 — D_0 Best Estimate

B.5 Ensuring Acceptable Edge Defect Densities

To ensure that an acceptable defect level is maintained for each of the oxide edges associated with a thin oxide (gate/drain, gate/field, plus others on more exotic devices), the same set of equations may be used if two minor changes are made:

1) In each instance in which an area is called for, a length is substituted. For example, rather than using A_c for the total area of the circuit, L_c would be used and would be a measure of, for example, the total gate to field oxide edge length of a circuit. Note that the units of measurement will be in cm rather than cm^2.
2) In each instance in which a defect density is called for, a defect per unit length value is substituted. This would have the dimensions of defects/cm rather than defects/cm^2. All calculations, strategies, procedures, and plots remain identical with that described in this annex.

Note that with this type of calculation, as well as with those in the remainder of this annex, only those defective oxides associated with the defect density to be measured should be considered. For example, when attempting to establish a defect density for polysilicon edge over thin oxide, only those structures that actually fail at that edge should be included. Other failures should be eliminated from the sample and replaced with an additional test structure. Failure analysis may sometimes be required to verify the defect location. Also, these procedures will not support the use of combined data from more than one test structure type as in B.3.

B.6 Examples for Use of Defect Density Curves

B.6.1 and B.6.2 give two examples of how a manufacturer can design a test plan using a minimum sample size and how to estimate defect density.

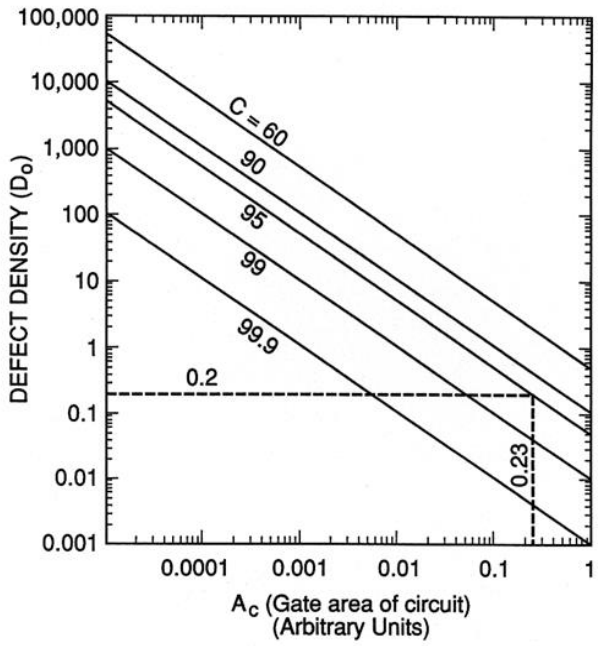
B.6.1 Example 1 — Designing a Test Plan

A manufacturer desires to have a 95% confidence level that each die shipped contains no defective oxides. Each die has a total active thin capacitor area of 0.23 cm^2. A test structure is available with an area of 0.013 cm^2. It is desired that the minimum number of structures be tested.

B.6.1.1 Determining Required Maximum Defect Density

The first step is to use Figure 18 to determine the defect density needed to provide the desired guarantee. Note that the units of the horizontal axis are not labeled. This is intentional, as any desired units can be used. Be aware, however, that defect density must from then on be calculated and expressed in terms of those same units. Consequently, by using an area measurement in square centimeters, it is necessary to use units of defects per square centimeter for all defect density terms. Figure 21 shows how Figure 18 would be used.

First, find the circuit area on the horizontal axis. Then, move vertically until intersecting the line for 95% confidence. Proceed to the vertical axis directly to the left of the intersection point and read off the value of defect density listed there. This is the defect density needed to satisfy the manufacturer's requirements. In this case, the value read from the figure is approximately 0.2 defects/cm^2.



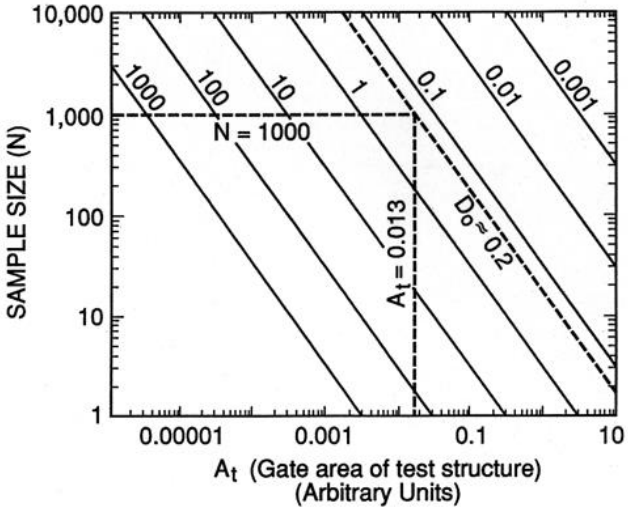
NOTE Illustration of how Figure 18 is used to determine the defect density needed to provide the desired guarantee. (See B.6.1.1)

Figure 21 — Determining D_o Needed to Provide a Desired Guarantee

B.6.1.2 Establishing a Sample Plan

In order to minimize the sample size required to prove a defect density of 0.2 defect/cm^2, a test should be developed that is passed only when no failures occur from among the sampled population. Allowing failures would require a larger sample size to prove the same defect density. With no failures allowed, Figure 19 may be used to determine the necessary sample size as shown in Figure 22.

First, find the area of the available test structure, 0.013 on the horizontal axis. Then, move vertically until the appropriate defect density line is intersected. In this case, the required defect density is 0.2 defect/cm^2, a value that is not explicitly plotted. Therefore, a "0.2 defect/cm^2" line will need to be interpolated between lines 0.1 and 1.0. Note that this should be a logarithmic interpolation. Following to the left from the intersection of this interpolated line gives a sample size of approximately 1000. Therefore, to make the desired claim using this test structure, 1000 structures must be tested and found to be defect free.



NOTE Illustration of how Figure 19 is used to determine the sample size N, given the area of the test structure A_t, and the defect density D_0. (See B.6.1.2)

Figure 22 — Determining Sample Size from D_o and Test Structure Area

B.6.1.3 Reducing the Required Sample Size

If the 1000 samples required in the above example is an unacceptably large sample size, several alternatives are available;

1) select a larger test structure,
2) provide a lower confidence level in your guarantee (e.g., 90% instead of 95%), or
3) use a smaller die with less total oxide area.

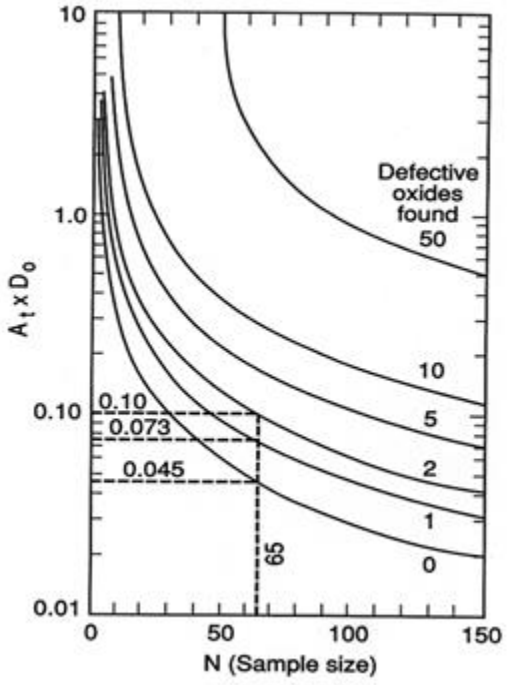
Note that in this example, a total test structure area of 1000 times 0.013 cm^2 was required to demonstrate a defect free oxide area of 0.23 cm^2. In other words, the ratio of area tested to area used was greater than 50. Although this number can vary greatly with confidence level and actual circuit area, it is not uncommon to have such a high ratio. This makes such a practice impracticable in a typical production environment.

B.6.2 Example 2 — Estimating Defect Density

Note that in the previous example, no information concerning defect density is obtained other than whether it is above or below the desired defect density. In order to estimate defect density, Figure 20 is useful. Suppose a plan is implemented using a sample size of 65 test structures of area 0.1 cm^2 and one defective oxide is found. Figure 23 demonstrates how Figure 20 can be used to show that this will ensure a defect density of approximately 0.73 defect/cm^2. This is obtained in the following manner:

1) Find the sample size (65) on the horizontal axis.
2) Move vertically until the "1 failure" line is intersected.
3) Then move left to the vertical axis and read the "A_tD_0 product." In this case, the product is 0.073.
4) To calculate defect density, simply divide the "A_tD_0 product" by the test structure area, A_t. Since the test structure area here is 0.1, the defect density that has been demonstrated is 0.73 defect/cm^2 (0.073/0.1).

If a number of failures other than one are found, then simply repeat the above procedure, except move across to the vertical axis when the appropriate "failure" line is reached. For example, no defective oxides would indicate an "A_tD_0 product" of 0.045, or a defect density of 0.45 defect/cm^2. Finding two defective oxides would indicate an "A_tD_0 product" of 0.1, or a defect density of 1.0 defects/cm^2. Again, if a failure line does not exist for the situation experienced, a logarithmic interpolation should be used.



NOTE Illustration of how Figure 20 is used to obtain the defect density D_0, given the sample size N, the number of failures, and the area of the thin oxide, A_t. (See section B.6.2)

Figure 23 — Determining D_o from Sample Size, Number of Failures and Oxide Area

Annex C (Informative) Fowler-Nordheim and Direct Tunneling Current

This annex is not a formal part of the test procedures but is included for informational purposes.

In an ideal oxide having a thickness greater than 5 nm, current density, J, should fit the Fowler-Nordheim equation:

$$J = AE^2 \exp\left(-\frac{B}{E}\right) \quad \text{(C.1)}$$

where A and B are constants related to the Si/SiO_2 electron effective mass and the barrier height, as determined from a gate current versus field measurement.

Typical values are in the order of:

$A = 1.6\ MA/(MV)^2$

$B = 222\ MV/cm$

Large deviations in the current density from the Fowler-Nordheim equation may indicate defective oxides.

For applied dielectric voltage less than the barrier-height ($F_B \sim 3.15$ eV for the Si/SiO_2 interface), the quantum mechanical tunneling mechanism switches from Fowler-Nordheim tunneling (triangle-barrier tunneling) as in equation C.1, to direct tunneling (so-called trapezoidal barrier tunneling). This mechanism becomes a dominant mechanism for ultra-thin dielectrics less than 4 nm at low voltages. A full quantum mechanical treatment of the direct tunneling mechanism is quite complex and beyond the scope of this document. Several approximated analytic solutions are available in the literature, but they involve various simplifying assumptions. In the direct-tunneling regime, a simple test of gate leakage at V_{use} is typically the only test used to determine whether or not a gate dielectric is defective before performing gate dielectric breakdown testing.

Standard Improvement Form

JEDEC JESD263

The purpose of this form is to provide the Technical Committees of JEDEC with input from the industry regarding usage of the subject standard. Individuals or companies are invited to submit comments to JEDEC. All comments will be collected and dispersed to the appropriate committee(s).

If you can provide input, please complete this form and return to:

JEDEC
Attn: Publications Department
3103 North 10th Street
Suite 240 South
Arlington, VA 22201-2107

Fax: 703.907.7583

1. I recommend changes to the following:

☐ Requirement, clause number _____

☐ Test method number _____ Clause number _____

The referenced clause number has proven to be:

☐ Unclear ☐ Too Rigid ☐ In Error

☐ Other _____

2. Recommendations for correction:

_____

_____

_____

_____

3. Other suggestions for document improvement:

_____

_____

_____

_____

Submitted by

Name: _____ Phone: _____

Company: _____ E-mail: _____

Address: _____

City/State/Zip: _____ Date: _____


JEDEC®