BREAKDOWN AND CHARGE TRAPPING

IN SILICON DIOXIDE FILMS ON SILICON.

Stephanie K. Haywood

A thesis submitted to the C.N.A.A.
in partial fulfilment of the requirements
for the degree of Doctor of Philosophy.

Middlesex Polytechnic, U.K.
Katholieke Universiteit Leuven, Belgium.

March 1986

ACKNOWLEDGEMENTS

The efforts of many people made it possible for the work described here to be carried out and several others assisted with the actual thesis production. I would like to thank everyone who helped for their invaluable contributions. In particular, my thanks to Prof. John Butcher of Middlesex Polytechnic Microelectronics Centre for introducing me to this field, for the opportunity to pursue this research in collaboration with the University of Leuven and for his advice regarding the thesis content and presentation. Thanks also, to my friends and colleagues at Middlesex Polytechnic (especially Moni Deb, Dr. Peter Revell and Dr. Kumar Das) for many stimulating discussions during the time I spent in London. Susan Partridge, Head of the Advanced MOS division at GEC Research Centre in Wembley provided samples and access to equipment in the early days of this work for which I am most grateful.

Special thanks to everyone at ESAT and IMEC in Leuven for helping to make the years I spent there happy and productive. Dr. Roger DeKeersmaecker provided endless scientific advice and support throughout this time. A number of errors were avoided and several improvements made as a result of his critical reading of this manuscript. Marc Heyns and Dr. Michiel Hillen were the source of much practical advice concerning experiments and also of many of the computer programs used in this work. Marc, especially, gave generously of his time to discuss results and ideas. I would like to thank both of them as well as my other colleagues from the 'oxides' group at Leuven, in particular Josiane Vandenbroek and Ivan Darakchiev, for their friendship and collaboration. Thanks, also, to Paul Heremans and Guido Groeseneken for help with the charge pumping measurements.

I am indebted to Mario Reybrouck for sample preparation, always done with great care and skill and similarly to Albert Debie who built several of the measurement systems and maintained them all in excellent working condition. Tom Buytaert and Marleen Tempels have also assisted in numerous ways over the past year. Thanks also to Georges Dubuisson for drawing a number of the figures, to Dr. Christopher Morley for proof reading the final manuscipt and to Mrs.M.Morley for last minute typing of the figure captions. Finally, I am grateful to my flat-mate Huang Qiuting for putting up with me during the traumatic writing-up stage and for being such a calming influence.

ABSTRACT

Several aspects of breakdown and charge trapping in silicon dioxide (SiO_2) have been studied. Firstly, the locations of charge trapped immediately prior to breakdown and of defects created under the high field stress were established using the photo I-V and avalanche injection techniques. Both positive and negative charge was found and in all cases charge build-up under high field conditions was at the interfaces, not in the bulk of the SiO_2. Electron trap creation occurred predominantly near to the non-injecting interface. Q_{bd}, the total charge which can be injected prior to breakdown, was examined under different current injection conditions. It was found to be strongly dependent on the duty cycle, the temperature and the gate electrode and to vary as $1/E_{ox}$, where E_{ox} is the average field across the oxide. The Weibull distribution was found to describe well the statistics of breakdown in both wearout and dielectric strength measurements. It was shown that the Weibull parameters a and b have the same values whichever of these methods is used to measure them. The breakdown mechanism is probably the same in both wearout and dielectric strength measurements, therefore.

Oxide degradation was also examined under the less severe conditions of bias-temperature stress (BTS). It was shown, using avalanche hole injection before and after negative BTS, that the positive charge generated during negative BTS is due to trapping of holes in intrinsic hole traps. This is accompanied by interface state generation across the whole band gap. The density of these states is linearly proportional to the number of holes trapped. On subsequent application of a positive BTS, the holes were all detrapped or neutralised. At this time a peak also appeared in the interface state density at ~0.2 eV above midgap. This may be due to a redistribution of previously generated states rather than to the creation of new states. The charge pumping technique was used to show that the peak is in fact due to interface states and not to lateral non-uniformities in the surface potential.

CONTENTS

CHAPTER ONE:
INTRODUCTION

Amorphous silicon dioxide (SiO_2) is widely used as an insulator in integrated circuits because of its high dielectric strength and its relative ease of formation within planar silicon processing technology. It also finds applications as a surface passivation layer and as a diffusion barrier during the doping of silicon. Thermally grown SiO_2 has the best characteristics for electrical isolation purposes and this is generally used in active areas such as for the gate dielectric of an EAROM (Electrically Alterable Read Only Memory) or of a MOSFET (Metal Oxide Semiconductor Field Effect Transistor) where relatively high electric fields are likely to be encountered.

Figure 1.1 shows a simplified diagram of an n-channel (enhancement mode) MOSFET illustrating the role of SiO_2 in the operation of the device. Inset is the simpler MOS capacitor stucture.

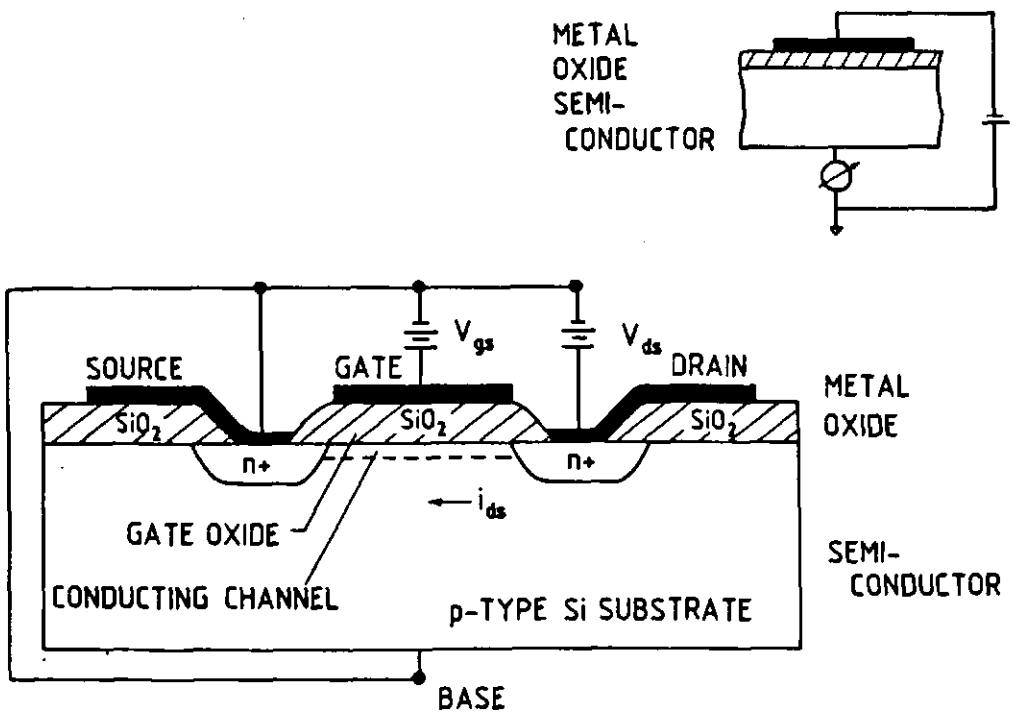


Figure 1.1 N-channel enhancement mode MOSFET and MOS capacitor.

The electric field at the $Si-SiO_2$ interface is controlled by the bias applied to the gate and the properties of the channel, a layer in the silicon adjacent to the interface, are thereby determined. At zero gate bias this region is p-type and does not conduct. No current will flow between source and drain although there is a voltage drop between them. However, when the voltage applied to the gate exceeds a certain threshold level, V_T, the channel region becomes inverted to n-type. Thus, there is now a conducting layer allowing current to flow between source and drain.

The number of transistors in an MOS integrated circuit (IC) has increased exponentially since 1960 (Fig.1.2, curve (a)). Although the rate of growth has slowed down over recent years, extrapolation from past trends still suggests that a one megabit memory chip (i.e. one million devices per IC) will be in production before 1990 (1). This is certainly a realistic estimate since such circuits have already been made at research level (2,3). The most important factor in achieving increased complexity is reduction of the minimum device dimensions (Fig.1.2, curve (b)) (4).

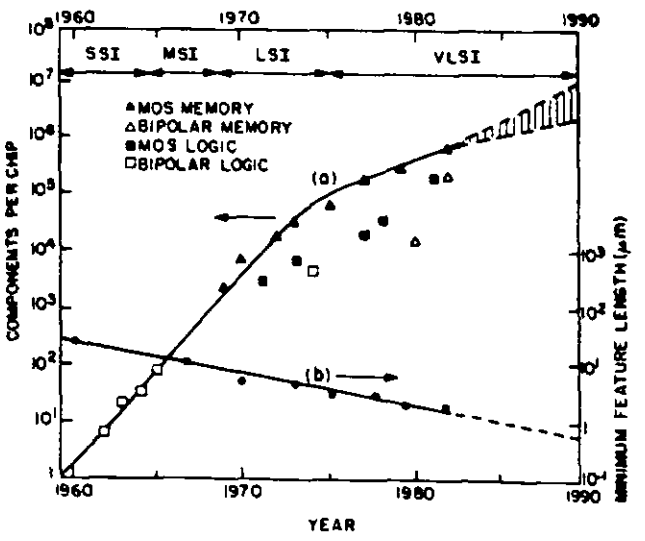


Figure 1.2 (a) Growth in number of components per chip (1) and
(b) Minimum device dimensions achieved since 1960 (4).

Scaling to smaller dimensions and the resulting higher level of integration which can be obtained reduces cost, gives higher operating speeds and lowers power consumption. For very large scale integration (VLSI), with $>10^5$ components per chip, high quality oxide layers of between 10 and 50 nm are required for the gate insulator. It is the properties of these thin thermally grown SiO_2 layers that will be considered in the present work.

Several consequences for the SiO_2 layer arise from device scaling. To begin with, some properties of the oxide may change as its thickness is reduced. For example, an advantage of thin oxides is that the maximum breakdown field increases with decreasing thickness (5,6). On the other hand defect densities (7,8) and interface trap densities (9) have been reported to increase thus reducing device reliability. The dielectric constant also seems to be different for very thin oxide layers (9,10). Furthermore, the voltage used for operation is often not scaled linearly with the oxide thickness so that the insulator may have to withstand higher applied fields. If this increase in field is severe it may increase the probability of electrical breakdown occurring during operation despite the reduced field sensitivity of thin oxides. There is also a greater chance of energetic electrons (or holes) being injected into the oxide from the channel and becoming trapped at defect sites or generating new defects in the oxide. More rapid degradation of the device characteristics may then result. Aspects of all the latter problems, breakdown, charge trapping and trap generation have been addressed in the course of this work.

Most measurements have been carried out on MOS capacitors rather than transistors. This simpler structure comprises the active region of the transistor (inset Fig.1.1) and requires fewer processing steps which is advantageous to the study of the intrinsic properties of the oxide layer.

1.1 The MOS capacitor.

The energy band diagram of an MOS capacitor with an aluminium gate electrode is shown in Fig.1.3. The physics of this device is discussed extensively in reference (11). Conduction is limited by the height of the energy barrier for excitation of electrons into the oxide conduction band, making the SiO_2 an excellent insulator. Since the barrier to hole injection is at least 1 eV higher than for electrons there will be negligible current due to hole conduction, especially as the hole mobility is also very low in SiO_2 (2×10^{-5} cm^2/Vs).

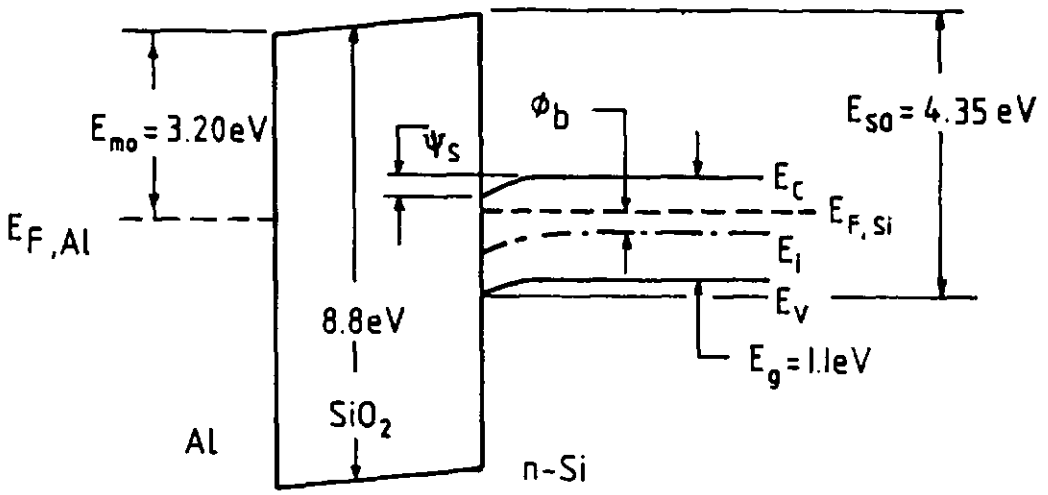


Figure 1.3 Energy band diagram of an MOS capacitor with an aluminium gate electrode (after (11).)

When the gate is connected to the substrate an internal electric field is maintained as a result of the difference in the work functions of aluminium and silicon. This work function difference, ϕ_{ms}, is given by:

$$q\phi_{ms} = E_{mo} - E_{so} + E_g/2 + q\phi_b$$

where E_{mo} and E_{so} are the energy barriers at the metal and silicon interfaces respectively. E_g is the bandgap and ϕ_b, the Fermi level in the silicon bulk. The sum of the band bending at the surface and the Fermi level, $\psi_s + \phi_b$, is the electrostatic potential at the surface or the surface potential, ϕ_s. If a gate voltage is applied to the device such that the band bending is zero, the so-called flatband condition is attained and the applied voltage is known as the flatband voltage (V_{fb}).

Although the energy barriers to conduction are high, once electrons enter the oxide conduction band they are quite mobile (~20cm^2/Vs at room temperature) and can flow under the influence of an electric field. At fields in excess of ~6MV/cm significant electron injection may occur due to Fowler-Nordheim (F-N) tunneling (12). Electrons from the metal or silicon conduction bands can tunnel through the triangular barrier at the interface which becomes narrower as the field is increased (Fig.1.4).

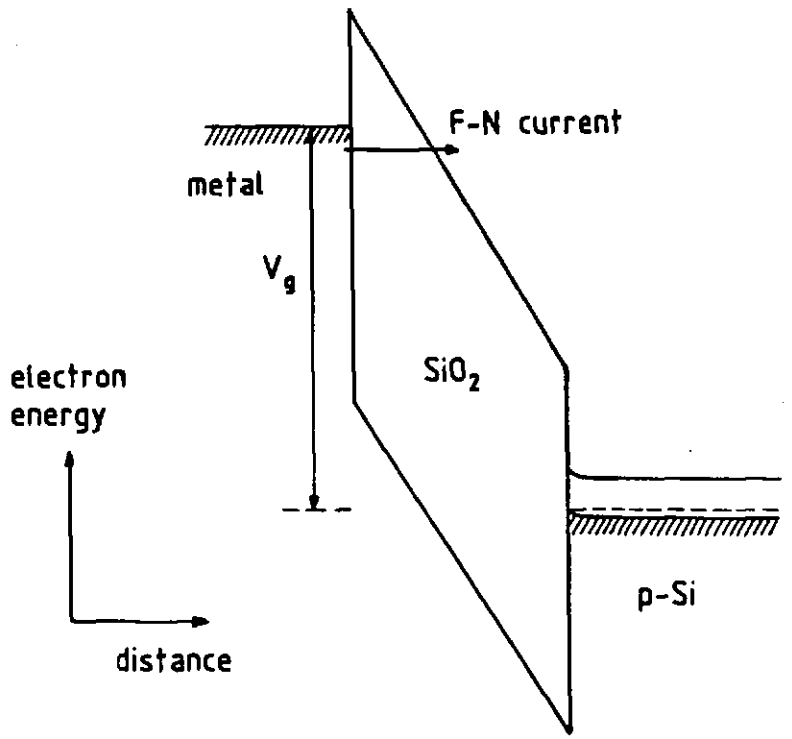


Figure 1.4 F-N tunneling into the SiO_2 conduction band.

Quite high currents may flow (up to -60A/cm^2) before the oxide breaks down. As the field is further increased however destructive breakdown eventually occurs. This is characterised by a sudden increase in current and loss of the insulating properties of the SiO_2 so that even when the applied field is lowered high leakage currents flow. A typical I-V curve measured on capacitor J on wafer JN1 and illustrating F-N tunneling and breakdown is displayed in Fig.1.5.

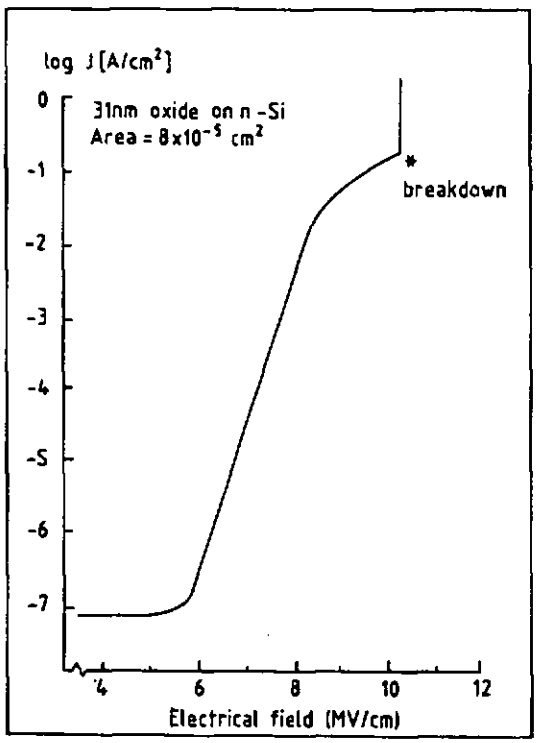


Figure 1.5 I-V curve for an MOS capacitor showing the displacement current, F-N tunneling current and breakdown.

1.2 Breakdown of SiO_2 layers.

Breakdown has largely been studied by means of two types of experiment. In a time dependent dielectric breakdown (TDDB) or wearout measurement, a constant voltage (13) or constant current (8) stress is applied and the time to breakdown measured. Typically a bimodal distribution is found (Fig.1.6). The early device failures are due to electrically active defects in the oxide which allow a

conducting pathway to form more readily on application of a bias. Constant voltage wearout measurements are frequently used to assess the reliability of SiO_2 layers under device working conditions. Extrapolation may be made from test conditions using elevated fields and temperatures to the normal operating situation (14).

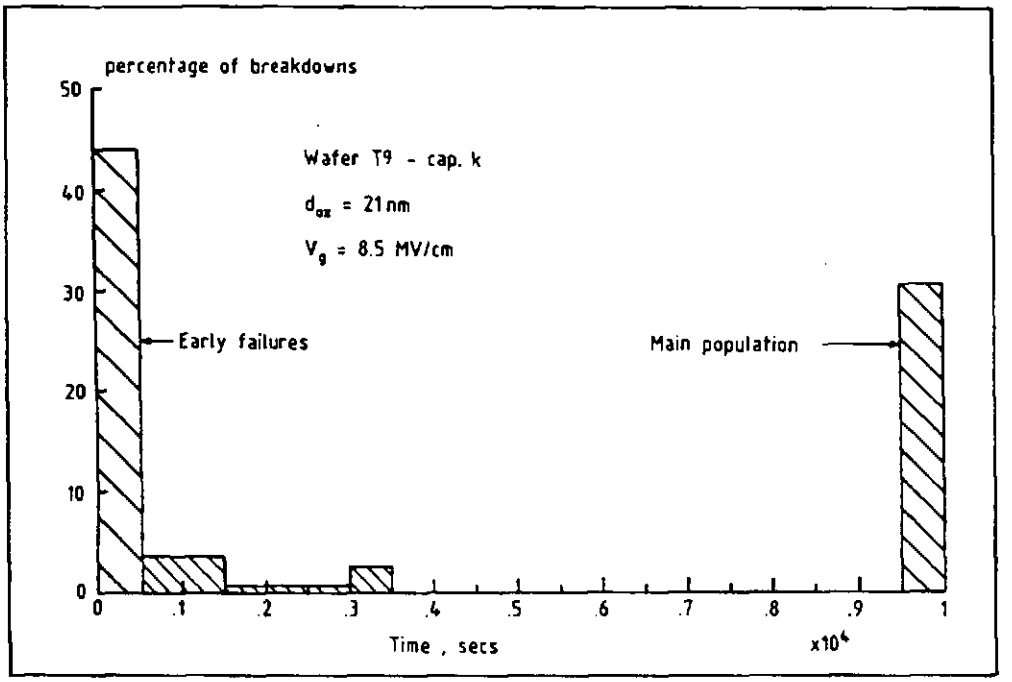


Figure 1.6 Example of a bimodal distribution typically observed for TDDB of MOS capacitors.

Another type of experiment commonly used to examine oxide integrity is the dielectric strength measurement. Here a fast voltage ramp is applied to each of a number of capacitors and the field at which breakdown occurs is measured. A statistically valid sample size, usually ≥100 devices, is taken and the results are presented in the form of a histogram. (This experiment is sometimes referred to as a 'breakdown' measurement. We shall however reserve the term breakdown for the actual destruction of the insulator which occurs in both wearout and dielectric strength measurements.) Breakdown is usually defined in a dielectric strength measurement as the exceeding

of a certain preset current. In order for this to be a valid definition, the current must be chosen carefully to ensure that irreversible failure has occurred i.e. that it is breakdown and not merely tunneling injection which is being observed. Three peaks are frequently observed in the resulting failure distribution (15,16). A typical histogram is shown in Fig.1.7 for cap. J on wafer H1.

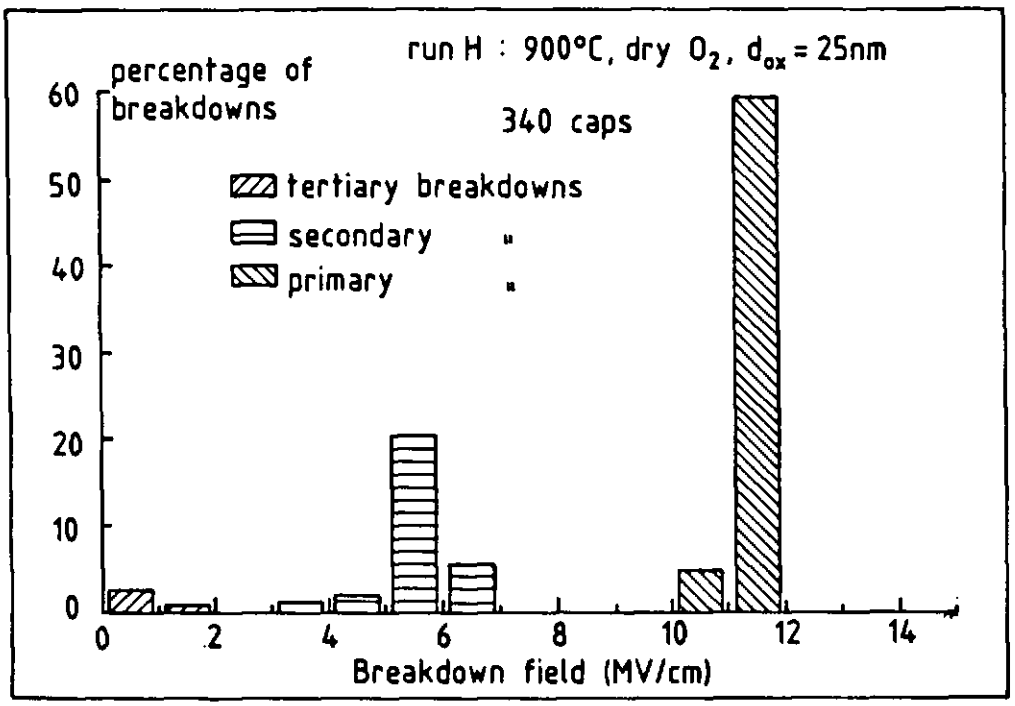


Figure 1.7 Breakdown histogram showing the three peaks commonly observed in a dielectric strength measurement.

The low field breakdowns (tertiary peak) have been attributed to gross structural defects such as pinholes. Medium field breakdowns (secondary peak) are most probably caused by smaller scale structural defects or impurities. The high field (primary) peak has frequently been assumed to represent intrinsic breakdown of the SiO_2. However, the slight effect of processing conditions on the position and width of this peak (17) and the increase in the maximum breakdown field (E_{bd}(max)) achieved over recent years (18), suggest that this point may not yet have been reached. Very small scale defects, perhaps

broken bonds or trapped charges, probably also play a role in causing high field breakdown and in determining the lifetime of a device under wearout conditions. Certainly TDDB has been shown not to be determined by the oxide alone but to be a property of the SiO_2-electrode system (19). Higher breakdown fields have also been achieved, at least at negative polarity, using the corona discharge technique rather than metal electrodes (20). The effect of processing and materials on breakdown and also its dependence on measurement parameters are discussed further in sections 3A.1 and 3A.2.

As device dimensions are scaled down and the SiO_2 thickness is reduced, the defect density (defined empirically in terms of those defects and impurities resulting in low and medium field breakdowns) tends to go up, causing yield problems. In part this problem can be solved by using extremely clean processing conditions. For example, use of a double-walled oxidation tube (DWO) removes most medium field breakdowns (21). This is thought to be because metallic impurities are carried away by the gas flow in the outer tube and thereby prevented from diffusing into the oxidation tube itself (22).

The primary peak moves to higher fields on reduction of the oxide thickness (23,24). The reason for this is not known and neither is the breakdown mechanism in this region well understood. Various models for intrinsic breakdown can be found in the literature. These fall essentially into two categories, those based on the build-up of positive charge, such as the impact ionisation model (25-29) and those based on electron trapping, such as the electron trap generation model of Harari (8,30) and the microplasma breakdown model of Budenstein (31) and Wolters (32,33). In section 3A.2.3 these will be considered in more detail, while in 3B.1 experimental evidence will be presented on the nature of the charge build-up occurring under the high oxide fields prevailing immediately prior to breakdown. Both positive and negative charge was observed. All charge trapping and defect generation occurred at or near to the interfaces. There was no evidence for bulk charge trapping under high field conditions.

Various experimental techniques were used for these measurements including high field constant current stress to simulate the pre-breakdown situation, avalanche electron injection (34) to fill oxide traps and internal photoemission (photo I-V) to examine the charge distribution in the oxide (35). These will be discussed in detail in section 2.3, while classical breakdown measurements, dielectric strength and TDDB are considered in 2.2.

The constant current stress technique offers certain advantages over the more commonly used fast voltage ramp and constant voltage wearout measurements. Firstly, it enables compensation for the varying internal oxide field, which changes as a result of charge trapping, and thereby allows the injecting field to be kept constant. Secondly, the total amount of injected charge prior to breakdown, Q_{bd}, can be simply calculated. This is a useful parameter to use in comparing the oxide quality of different samples, providing the injection is carried out at the same current density. It will be shown in 3B.2 that, in a constant current measurement, Q_{bd} is not independent of the injection conditions, such as current density and duty cycle, as has been previously suggested (8,36,37). Unlike dielectric strength measurements, wearout also allows examination of the events leading up to breakdown in 'slow motion.' This does presuppose however, that breakdown itself takes place by the same mechanism in a fast voltage ramp measurement as it does under wearout conditions. There is some evidence in the literature that this is indeed the case (14,38,39). To examine this point further, the statistical relationship between dielectric strength and time-to-breakdown in a constant voltage wearout measurement was also studied (3B.3).

1.3 Charged defects in SiO_2 layers.

Defects may be present in as-grown oxides or be generated subsequently under various conditions. Apart from any role they may play in enhancing breakdown these defects may have deleterious effects

on device operation should they become charged. The Deal notation and classification scheme depicted in Figure 1.8 will be used to define operationally the types of defect which may be found in the MOS system (40). Q_m represents mobile ionic charge, such as alkali ions, which can migrate through the oxide under bias (41), especially if elevated temperatures are used. In particular, sodium is easily introduced from the environment. Contamination of this kind was a severe problem in the early days of silicon technology but it has been largely overcome by the use of very clean processing conditions.

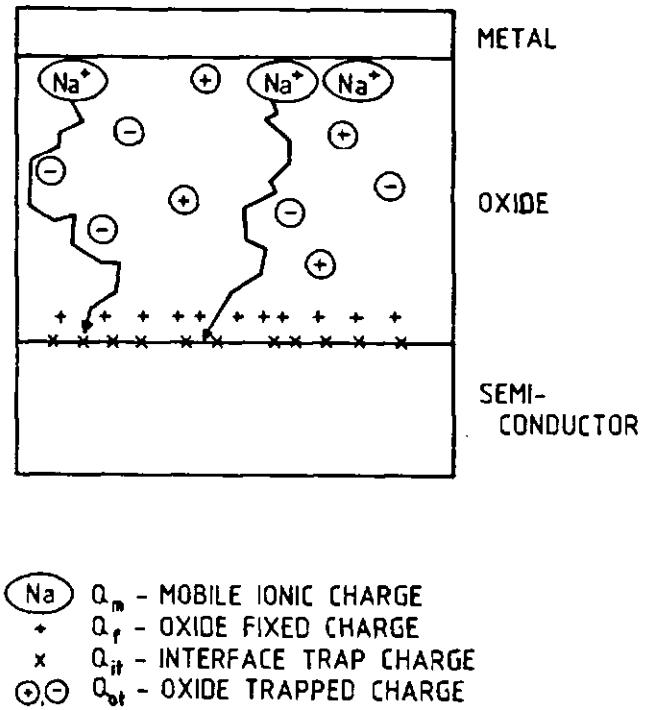


Figure 1.8 Deal notation and classification scheme for charged defects in MOS structures (taken from ref. (42)).

Examples of defects whose charge state may be varied are both fast and slow interface states. The former may be positively or negatively charged or neutral. They are often referred to simply as 'interface states' but they have also been called 'fast states', 'interface traps' or 'surface states'. In Deal's system they are represented by Q_{it}. They are located physically at the interface and energetically within the silicon bandgap. Consequently, the

charge state of these defects depends on the position of the Fermi level at the interface and hence on the applied voltage. The density of interface traps is usually specified with respect to unit energy within the silicon bandgap, that is in /cm^{2}eV for which the symbol D_{it} is used. (N_{it} is used for $\int D_{it}.dE$.)

Interface states are present in as-oxidised samples with a density of the order of 10^{11}-10^{12} /cm^{2}eV (42). Their density is reduced by a post-metallization anneal (PMA) generally performed in forming gas. After this treatment the trap density displays a minimum at midgap with an increase towards both band edges giving a U-shaped distribution. The lowest reported midgap values are around 5 x 10^8 to 10^9 /cm^{2}eV (43,44). Using electron spin resonance (ESR) Caplan et al. (45) have demonstrated these interface traps to be due to $Si_3{\equiv}Si\cdot$ species or so-called dangling bonds at the interface. As well as being present in as-processed oxides, interface traps are also produced by various forms of electrical stress e.g. avalanche injection of electrons (46) or holes (47), high field stress (48) or bias-temperature stress (BTS) (48,49) and also by ionising radiation (48,50,51,52). Methods of measuring interface state densities are discussed in section 2.4.

'Slow-states' are also located at or very near to the interface but they respond more slowly to an applied bias. They have been defined somewhat arbitrarily as those not responding to a 1 MHz high frequency C-V sweep at 1 V/s (53). These defects may be either positively charged or neutral and they were first seen in avalanche electron injection experiments where they are responsible for the 'turn-around' effect (54,55). (The flatband voltage shift is initially positive during avalanche electron injection, indicative of electron trapping in the oxide. After injection has continued for a while, however, the shift becomes negative showing the presence of positive oxide charge. This is known as the 'turn-around' effect.) The generation of slow states has now also been observed during injection by F-N tunneling (21) and avalanche injection of holes (56).

Slow-states are converted to a neutral charge state by the application of moderate positive biases or elevated temperatures and positively charged again by negative applied biases.

Fixed oxide charge, Q_f, is another interface related defect, always positive in sign. It appears to be physically further from the interface and/or energetically outside the silicon bandgap, since its occupancy is not a function of applied voltage. Immediately after oxidation, Q_f is present with a density dependent on the oxidation temperature. Its density is usually of the order of $\sim 10^{12}$ /cm^2 and may by reduced to 10^{10}-10^{-11} /cm^2 by a post-oxidation anneal (POA) in an inert ambient such as nitrogen (N_2) or argon (Ar) (42). D_{it}(init), the initial as-processed interface state density, and Q_f respond in the same way to oxidation conditions and some but not all annealing treatments (57), implying a connection between them as regards physical origin.

Charge of either sign may be distributed throughout the oxide and is denoted by Q_{ot} (oxide trapped charge). This oxide charge may be introduced during fabrication, as a result of irradiation and by the same experimental techniques which create interface states: avalanche injection of electrons or holes, high field stress or BTS.

1.4 The Negative Bias Instability

Sodium contamination results in an instability in devices under positive bias conditions where drift of Na^+ ions to the Si-SiO_2 interface results in a positive shift of the threshold voltage, V_T. In devices free of mobile ions, another phenomenon is sometimes observed known as the negative bias instability. Prolonged operation of MOSFETs under negative gate bias, causes both interface states and positive oxide charge (denoted here as $Q_{ot}(+)$) to be generated close to the Si-SiO_2 interface resulting in a negative shift of V_T (58). This can threaten the reliability of MOSFETs, especially of p-channel devices (59). As well as causing a V_T shift, drain junction

avalanche breakdown is also affected, being reduced for p-channel and increased for n-channel transistors. Gain or transconductance is reduced by both oxide charge and interface states. A larger change in gate voltage is required to produce a given change in channel conductance when D_{it} is large. Also the gain depends on the effective mobility of free carriers in the channel and this is reduced by scattering at charged centres. Drain junction leakage current and flicker noise are also increased by interface states.

As device dimensions are reduced, interface effects such as the negative bias instability become relatively more important. Low-cost plastic packaging which is being used increasingly has also been found to enhance the negative bias instability and has thus generated renewed interest in this long-standing problem (60,61). It is thought that the diffusion of water from the packaging or from phosphosilicate glass (PSG) passivation layers may be the cause of this enhanced degradation.

Measuring the shift in the high frequency capacitance-voltage (C-V) curve on an MOS capacitor is a convenient means of studying this instability and application of the bias at elevated temperatures speeds up the ageing process so that experiments can be conducted on a more manageable time-scale. This technique is in fact similar to the wearout measurements discussed in 1.2. In BTS, rather than applying a high bias at room temperature, a more moderate bias is used at elevated temperature. The point of breakdown is not normally reached under the less severe BTS conditions, but both wearout and BTS are essentially ageing techniques.

Figure 1.9 shows typical high frequency C-V curves measured following various BTS treatments. After negative BTS (curve (B)) there is a negative shift of the C-V curve due to $Q_{ot}(+)$ and a distortion relative to the initial curve (curve (A)) due to Q_{it}. The hysteresis is indicative of some generation of slow-states. $Q_{ot}(+)$ shows many similarities to Q_f already present in the oxide.

For example, for a given set of BTS conditions the magnitude of $Q_{ot}(+)$ generated is proportional to Q_f (58). It has been suggested that the two types of charge may be due to the same or similar defects. However, they do show some different behaviour, for example, with respect to low temperature annealing (62). The number of interface states generated during negative BTS appears to be proportional to the initial interface state density, $D_{it}(init)$ (64). Hence these defects are probably also very similar to each other.

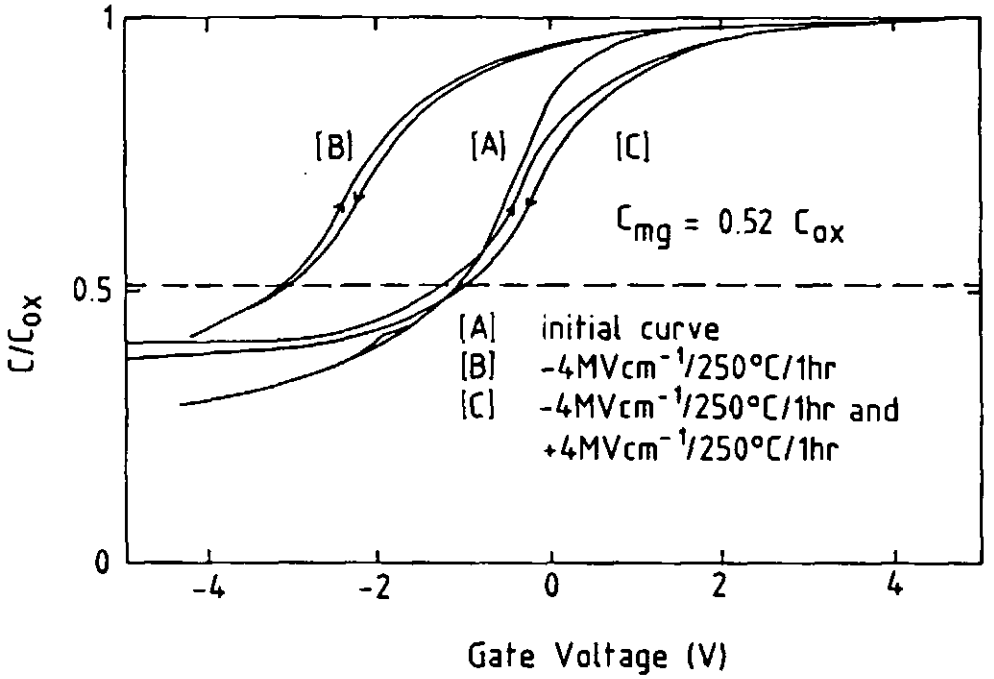


Figure 1.9 Shifts in the high frequency C-V curve after various bias-temperature stresses.

Applying a positive BTS after the negative one (curve (C)) returns the C-V curve to more positive voltages. The distortion of the curve is greater, however, indicating considerable interface degradation. Positive BTS alone results in at most a very small shift to positive voltages (63), indicative of electron trapping. Some interface state generation also occurs.

Despite extensive studies of both the kinetics and the processing dependence of the negative bias instability (considered in 4A.1 and 4A.2 respectively), the mechanism of positive charge generation is still uncertain. Various models have been proposed, the most important being hole trapping (65) or electron emission from traps in the oxide (66), the breaking of weak bonds associated with the already partially ionised silicon atoms near the interface (58) and the formation and migration of oxygen vacancies (68). These are discussed in more detail in 4A.3. To try to shed more light on these models, the avalanche hole injection technique (outlined in 2.3.3) was used before and after various stresses in an attempt to elucidate any role hole traps might play in positive charge generation during negative BTS. $Q_{ot}(+)$ was in fact found to be due to filling of intrinsic hole traps i.e traps already present in the oxide before it was subjected to BTS (4B.2). Interface state generation and its relationship to hole trapping was also examined (4B.3) and a kinetic study of the voltage shift of the C-V curve during BTS was carried out (4B.1).

CHAPTER TWO:
EXPERIMENTAL TECHNIQUES AND ANALYSIS

2.1 Sample Preparation.

Oxides were grown at 850-1000°C to thicknesses of between 10 and 60 nm on both n- and p- type <100> silicon wafers. 10^{15} or 10^{17} /cm^3 doping was used, the heavily doped material being to ensure laterally homogeneous injection during avalanche injection measurements. Most wafers were processed in a double-walled oxidation tube (Fig.2.1) to reduce the number of medium field breakdowns (21). The oxidation ambient was usually dry oxygen (O_2) with O_2 plus <1% trichloroethane (C33) in the outer tube. In most cases, a short POA (about 10 mins.) was carried out in N_2.

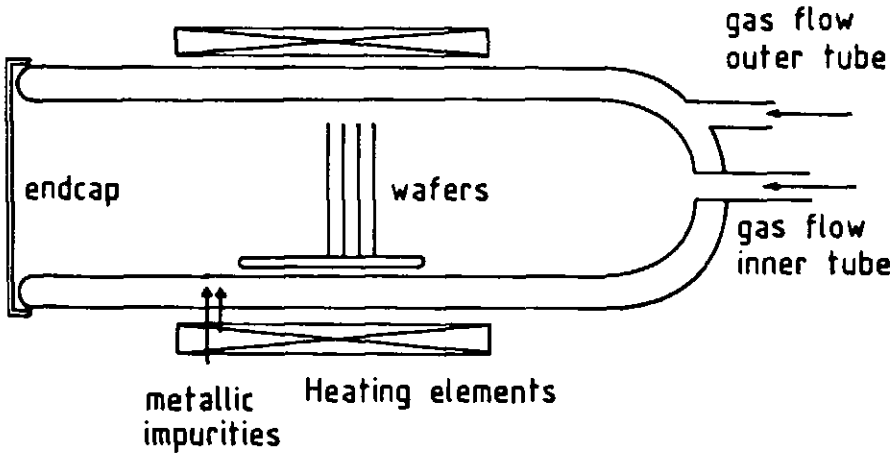


Figure 2.1 Cross section of a double-walled oxidation tube. Metal impurities are removed by the gas flow in the outer tube.

Key processing details for each of the wafers used in this work are shown in Table 2.1 on the following three pages. Aluminium deposition was by d.c. magnetron sputtering or flash evaporation from a tungsten boat. After definition of the electrodes by standard wet lithography, a post-metallisation anneal was carried out in forming gas at 435°C

WAFER CODE	SUBSTRATE DOPING	OXIDATION CONDITIONS	POA	ELECTRODE MATERIAL
232X	p/1E15	52 nm/900°C/2%HCl (GEC)	15" N_2	1μ Al/Si mag. sputt.
H1	n/1E15	25 nm/900°C/dry O_2	10" N_2	1μ Al mag. sputt.
JN1	n/1E15	31 nm/850°C/dry O_2	--	1μ Al mag. sputt.
DP2	p/1E15	35 nm/900°C/dry O_2	10" N_2	1μ Al mag. sputt.
L9	n/1E15	39 nm/900°C/dry O_2 (DWO)	10" N_2	1μ Al mag. sputt.
MN2	n/1E15	39 nm/900°C/dry O_2	10" N_2	1μ Al mag. sputt.
MN6	n/1E17	39 nm/900°C/dry O_2	10" N_2	1μ Al/Si mag. sputt.
MP2	p/1E15	39 nm/900°C/dry O_2	10" N_2	1μ Al/Si mag. sputt.
MP6	p/1E17	39 nm/900°C/dry O_2	10" N_2	1μ Al/Si mag. sputt.
P4A3	n/1E15	38 nm/900°C/dry O_2 (DWO)	10" N_2	Poly 440 nm I/I P 1E16
P512	n/1E15	29 nm/900°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	Poly 440 nm P diffusion
S02	n/1E15	31 nm/900°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	Poly 440 nm I/I P 1E16
U3	n/1E15	59 nm/1000°C/dry O_2 (DWO:0.2% C33 O.T.)	--	1μ Al/Si mag. sputt.
U7	n/1E15	39 nm/900°C/dry O_2 (DWO:0.2% C33 O.T.)	--	1μ Al/Si mag. sputt.
U8	n/1E15	39 nm/900°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	1μ Al/Si mag. sputt.
U9	n/1E15	27 nm/900°C/dry O_2 (DWO:0.2% C33 O.T.)	--	1μ Al/Si mag. sputt.

Table 2.2 Summary of wafer processing details (continued overleaf).

U10	n/1E15	28 nm/900°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	1μ Al/Si mag. sputt.
U11	n/1E15	12 nm/900°C/dry O_2 (DWO:0.2% C33 O.T.)	--	1μ Al/Si mag. sputt.
U12	n/1E15	12 nm/900°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	1μ Al/Si mag. sputt.
13N	n/1E15	25 nm/900°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	Poly 450 nm B diffusion
A11	n/1E15	45 nm/900°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	15 nm Al mag. sputt.
B6	n/1E15	23 nm/900°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	Poly 450 nm P diffusion
CP1	n/1E15	23 nm/900°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	Poly 450 nm P diffusion
F6	n/1E15	54 nm/1025°C/dry O_2 (DWO:0.2% C33 O.T.)	15" N_2	Poly 450 nm P diffusion
ST1	n/1E15	28 nm/900°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	15 nm Al mag. sputt.
ST2	n/1E17	28 nm/900°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	1μ Al/Si mag. sputt.
S1	n/1E17	51 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	1μ Al mag. sputt.
S2	n/1E17	51 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	500 nm Al mag. sputt.
S3	n/1E17	51 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	250 nm Al mag. sputt.
S4	n/1E17	51 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	125 nm Al mag. sputt.
S5	n/1E17	51 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	50 nm Al mag. sputt.
S6	n/1E17	51 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	15 nm Al mag. sputt.

Table 2.2 (cont'd) Summary of wafer processing details.

T1	n/1E15	80 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	15 nm Al mag. sputt.
T2	n/1E15	80 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	60 nm Al mag. sputt.
T3	n/1E15	80 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	1μ Al mag. sputt.
T6	n/1E15	12 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	1μ Al/Si mag. sputt.
T9	n/1E15	21 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	1μ Al/Si mag. sputt.
T12	n/1E15	49 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	1μ Al/Si mag. sputt.
R1	n/1E17	58 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	Poly 450 nm + 1μ sputt.
R2	n/1E17	58 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	Poly 450 nm + 425 nm sputt.
R3	n/1E17	58 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	Poly 450 nm + 100 nm sputt.
R4	n/1E17	58 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	Poly 450 nm + 850 nm flash
R5	n/1E17	58 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	Poly 450 nm + 425 nm flash
R6	n/1E17	58 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	Poly 450 nm + 100 nm flash
R7	n/1E17	58 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	Poly 450 nm + no metal
R8	n/1E17	58 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	850 nm Al mag. sputt.
R10	n/1E17	58 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	850 nm Al flash evap.
R12	n/1E17	58 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	425 nm Al flash evap.
R13	n/1E17	58 nm/950°C/dry O_2 (DWO:0.2% C33 O.T.)	10" N_2	100 nm Al flash evap.

Table 2.2 (cont'd) Summary of wafer processing details.

for 20 mins. Polysilicon gates were formed by LPCVD of polysilicon (450 nm) which was then thermally doped with $POCl_3$ or ion-implanted with boron or phosphorus. Aluminium was then deposited on top of the polysilicon to ensure good contact. Capacitors of various sizes were defined using one of the two sets of masks illustrated in Fig.2.2. The capacitor dimensions are shown in Table 2.2 for the case of thin oxide only and when using a thick field oxide. The capacitor area is reduced slightly in the latter case by the area of the central contact pad which covers thick oxide rather than thin oxide. Most of the devices measured were processed at ESAT, K.U.Leuven. However, wafer 232X in Table 2.1 was part of a batch made at GEC Hirst Research Centre, Wembley.

Capacitor Code	Thin Oxide Only Area (cm^2)	Thin Oxide + Field Oxide Area (cm^2)
A	4.42×10^{-3}	4.11×10^{-2}
B	4.91×10^{-4}	3.74×10^{-4}
C	1.26×10^{-3}	1.26×10^{-3}
D	1.66×10^{-3}	1.51×10^{-3}
E	1.15×10^{-2}	1.11×10^{-2}
F	1.08×10^{-3}	9.39×10^{-4}
G	4.41×10^{-3}	4.12×10^{-3}
H	7.54×10^{-4}	6.28×10^{-4}
I	1.26×10^{-3}	(Field oxide only)
J	7.85×10^{-5}	7.85×10^{-5}
K	3.14×10^{-4}	3.14×10^{-4}
L	4.82×10^{-4}	4.82×10^{-4}
M	1.96×10^{-5}	1.96×10^{-5}

Table 2.2 Dimensions of capacitors on masks 1 and 2 in Figure 2.2.

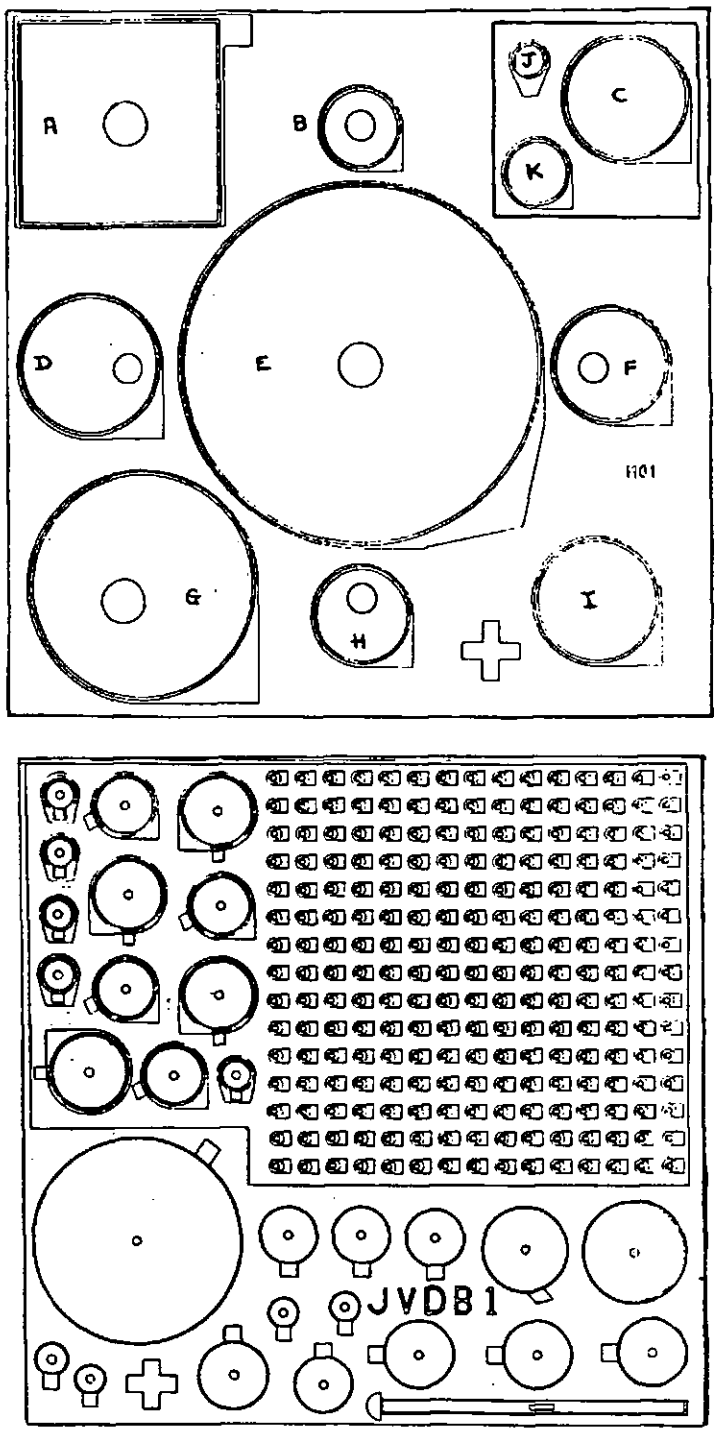


Figure 2.2 Photolithography masks used to define caps. of various areas, with and without a field oxide. Mask 2 has caps. B,F,G,H,J and K of mask 1 plus the rectangular cap., L and an array of small caps., M.

2.2 Breakdown measurements.

Constant voltage wearout and fast voltage ramp (dielectric strength) breakdown measurements were both carried out on an automatic measurement system. This is controlled by an HP 9835 desk top computer via an IEEE bus and includes a Karl Süss wafer probe station and a multi-probe breakdown detection system built at ESAT. Before describing this system in more detail some of the problems associated with breakdown measurements will be considered.

2.2.1 Dielectric strength.

A fast voltage ramp is generally applied to a capacitor in order to measure the dielectric strength of the insulator i.e. the field at which it would break down on instantaneous application of the voltage. Ideally an infinitely fast ramp would be used so that no wearout could occur during the measurement and the true time-zero breakdown would be measured. In practice, however, the need for a reliable means of breakdown detection limits the ramp rate which can be used. Usually, breakdown is defined experimentally as the exceeding of a preset current. The voltage ramp must therefore cause only a small displacement current relative to this set current. The response time of the measuring circuit and the accuracy required in measurement of the breakdown voltage must also be considered.

The current chosen to define breakdown, I_{tr}, is an extremely important parameter (69). In the high field part of the breakdown distribution very large currents may flow prior to catastrophic failure (up to ~60 A/cm^2 was measured on 8 x 10^{-5} cm^2 samples). If the current at breakdown, I_{BD}, is high a large trigger current must be chosen to ensure the capacitor is actually broken down. Under these conditions a relatively large voltage drop across the substrate (and any resistance in the measurement circuit) has to be taken into account. This is V_{Si} in Fig.2.3. V_{Si} will be especially

significant for thin oxides and low substrate dopings. For high field breakdowns it is simply a matter of subtracting V_{Si} from the measured breakdown voltage. All low field breakdowns occurring below $I_{tr}.R_{Si}$ (= V_{BD}^*) will be measured as V_{BD}^*, however (see Fig.2.3). Hence it is not always possible to observe high and low field breakdowns simultaneously. The recently proposed technique (7D) in which a high voltage pulse is followed by a lower voltage leakage test avoids the problem of choosing an appropriate set current but the parasitic voltage drop must still be taken into account. Harari (8) applied a high constant current stress to his samples and measured the time-to-breakdown, t_{bd} and also the applied voltage. Extrapolation of a plot of t_{bd} versus I to zero time gave I_{BD}. V_{BD} could be calculated from this technique via the equation for F-N tunneling.

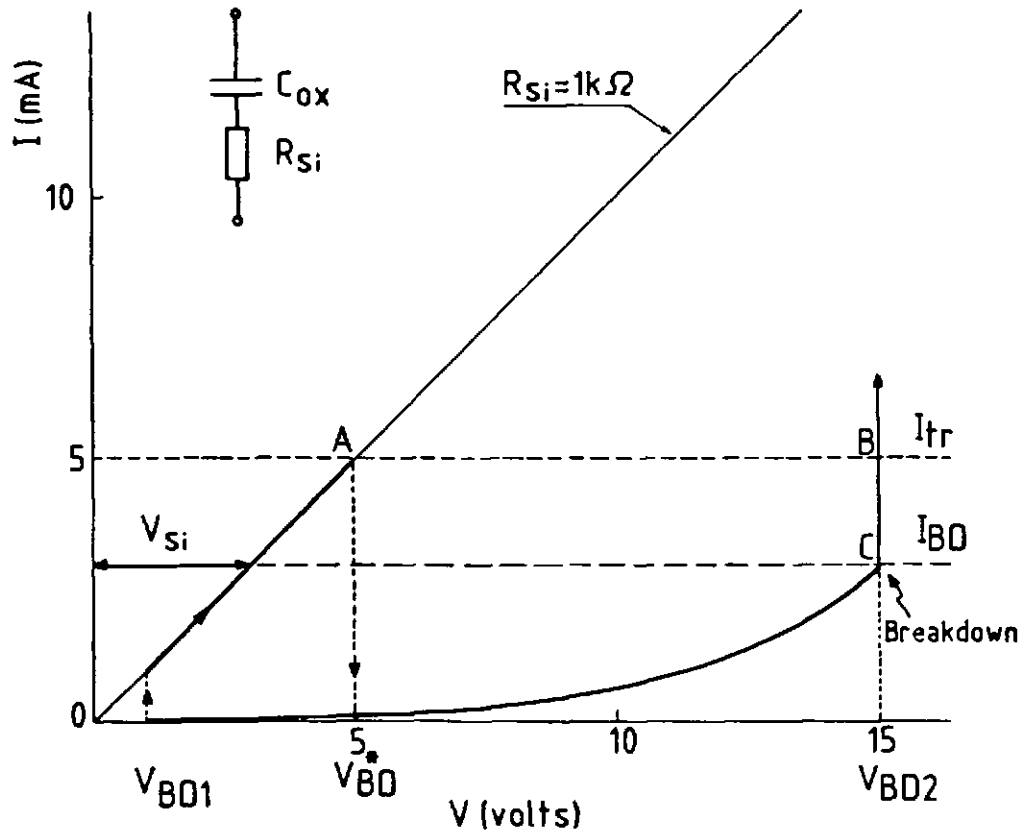


Figure 2.3 Effect of series resistance and trigger current on the breakdown voltage in a dielectric strength measurement.

2.2.2 Time dependent dielectric breakdown.

Time dependent dielectric breakdown (TDDB) or wearout measurements may be conducted by applying either a constant current or constant voltage stress to the sample and monitoring the time at which the sample breaks down. These measurements are by their nature time consuming and it is advantageous to be able to examine several devices simultaneously. The multi-probe system described below allows 36 capacitors to be stressed at the same time under constant voltage conditions. Constant current measurements were only conducted on individual devices. This could be done in either of two ways. On the avalanche injection system (described in section 2.3.5) a variable d.c. voltage could be used in place of the sawtooth waveform. Alternatively, a Keithley 220 programmable constant current source was available for use in conjunction with an HP 3456A digital voltmeter. These devices were also controlled by the HP 9835A computer and this system allowed times-to-breakdown as low as 3 ms to be measured and hence high current stresses to be applied. I-V curves could also be measured using this configuration.

2.2.3 Multi-probe breakdown detection system.

A block diagram of this system is shown in Fig.2.4. The voltage source (-100V to +100V) has a measurement accuracy of 25 mV and five possible sweep rates: 2.5, 5, 10, 25, 50 and 100 V/s. The reference current can be chosen over the range from 0.08 to 5 mA. Time can be measured up to 174 min with an accuracy of 0.01 seconds. Therefore, both wearout and fast voltage ramp measurements may be performed on this system.

Each measurement consists of several cycles during which all 36 capacitors are tested. For dielectric strength measurements the value of the applied voltage is determined by a time base, a 12-bit counter and a D/A converter. The voltage is incremented by 25 mV at each clock pulse. The length of this pulse is variable from 250 to 10 µs.

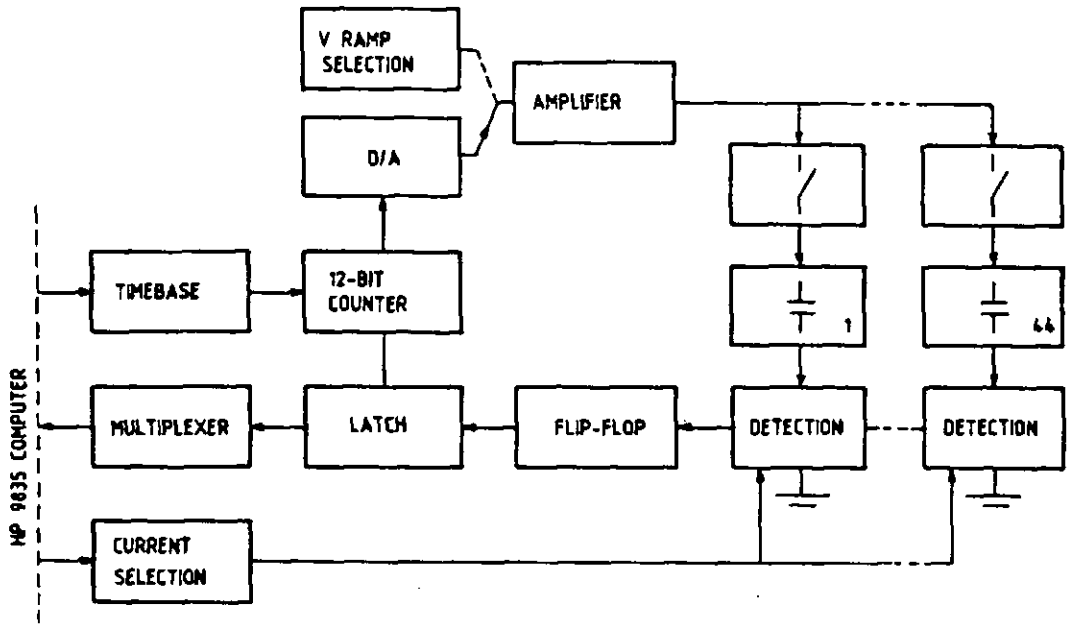


Figure 2.4 Block diagram of the multi-probe breakdown detection system.

Thus, sweep rates from 2.5 V/s to 100 V/s can be obtained. When one or more capacitors exceed the reference current during a voltage step a flip-flop corresponding to that capacitor is set and an interrupt is sent to the microprocessor. At the end of each voltage step the applied voltage and the state of all the flip-flops are read into the microprocessor. Broken-down oxides are immediately disconnected from the voltage source. Before the voltage is incremented again all the flip-flops are reset, the data is processed by the microprocessor and the next measurement cycle begins.

For wearout measurements a constant voltage is applied to the capacitors and on breakdown the time instead of the voltage is read in. The 12-bit counter is extended to a 20-bit counter for these measurements to increase the maximum time which can be registered to 174 minutes.

2.3 Charge injection and sensing techniques.

Three techniques were available to inject charge into MOS capacitors. The first of these was simply the application of a high field to the oxide which results in current flow by F-N tunneling. Tunneling injection has been used for charge trapping studies (71,72,73) but the requirement for high fields to inject relatively low current densities is a disadvantage for this purpose. Field ionisation of trapped charge and breakdown of the SiO_2 are favoured under these conditions. In this work high field stressing was used primarily to simulate the conditions existing immediately prior to breakdown of the insulator. Charge was injected under constant current conditions, the voltage being adjusted to compensate for carrier trapping in the oxide. The charge which could be passed prior to breakdown, Q_{bd}, was also measured in this way.

Avalanche injection of both electrons (34,74) and holes (75) was used to charge traps in the oxide. Relatively large current densities can be injected into the SiO_2 by this method at moderate oxide fields ($\leq$4 MV/cm) enabling the detection of traps with small capture probabilities. The low fields used also limit further damage to the oxide and make detrapping of charge less likely.

The third charge injection technique used was internal photoemission or photo I-V (76,77). This allows the injection of low current densities at low to moderate fields and the I-V curve obtained is very sensitive to the charge distribution in the oxide. It was used to determine the centroid and density of trapped oxide charge (35).

These three techniques and the experimental set-up used to conduct the measurements are described in greater detail in the following sections.

2.3.1 Fowler-Nordheim tunneling.

On application of fields above ~6MV/cm substantial current begins to flow through an MOS capacitor. From the dependence of the current density on the applied field ($\log (J/E^2) \propto 1/E$) conduction would seem to be due to F-N tunneling of electrons through the triangular barrier at the interface into the oxide conduction band (12). The complete expression for the tunneling current is:

$$J = (q^2E^2/8\pi h\Phi)\exp[-4(2m^*)^{0.5}\phi^{1.5}/3q\hbar E] \qquad (2.1)$$

where h is Planck's constant, $\hbar$ is $h/2\pi$, q the charge on the electron, E the electric field, Φ the barrier height and m^* the effective mass of the electron (usually taken as 0.5 m, where m is the free electron mass).

If the effect of the image force on the barrier is taken into account, two correction factors t(y) and v(y) are introduced into the Fowler-Nordheim equation. These are tabulated integrals (78) dependent on the normalised image-force barrier lowering, y, where:

$$y = (1/\Phi)(q^3E/4\pi\varepsilon_{ox})^{0.5} \qquad (2.2)$$

This correction causes an essentially parallel shift of the theoretical F-N plot to higher values of J/E^2.

Although tunneling is itself temperature independent the number of electrons incident on the barrier depends on the temperature. This introduces a term $\pi ckT/\sin(\pi ckT)$ where $c = 2(2m^*\phi)^{0.5}t(y)/q\hbar E$ into the Fowler-Nordheim equation. The F-N plot is still close to a straight line but with a slightly smaller slope. The tunneling equation including correction factors then becomes:

$$J = (q^3E^2/8\pi h\Phi)[1/t^2(y)][\pi ckT/\sin(\pi ckT)] \qquad (2.3)$$
$$\times \quad \exp\{-[4(2m^*)^{0.5}\phi^{1.5}/3q\hbar E].v(y)\}$$

As well as using the F-N tunneling phenomenon for constant current high field stress measurements, ramp I-V curves were also measured. This could be done either by using a ramp voltage source (Sweepy 1, built at ESAT) and logarithmic picoammeter (Keithley 26000) with a Houston 2000 chart recorder or by means of the Keithley 220 programmable current source and HP 3456A digital voltmeter previously described.

2.3.2 Avalanche electron injection.

To cause avalanche injection of minority carriers, a periodic waveform is used (sinusoidal, square or sawtooth) to pulse the substrate into deep depletion (34,74). Carriers generated in the depletion layer reach sufficient energies for impact ionisation to occur. Some of the hot carriers thus generated have enough energy to surmount the interfacial barrier and enter the SiO_2. Electrons are injected when the substrate is p-type and holes when it is n-type. In this manner charge is injected over part of each pulse cycle. Charge trapping during injection means that to maintain a constant current a feedback circuit must be used to regulate the amplitude of the avalanching pulses. Avalanche injection can be performed only over a limited doping range. Above 10^{18} /cm^3 dopant atoms interband tunneling rather than avalanche breakdown occurs. Below ~5 x 10^{16} /cm^3, depending on the oxide thickness, edge breakdown occurs so that injection is no longer homogeneous (79).

During avalanche injection the flatband voltage shift (ΔV_{fb}) or the midgap voltage shift (ΔV_{mg}) is usually measured as a function of time (and hence of injected charge). Assuming first order trapping kinetics as proposed by Ning and Yu (80), these data can be converted to a set of discrete traps with capture cross sections σ_i and effective trap densities, N_i according to the following equation:

$$\Delta V_{fb} = \sum_{i=1}^{n} qN_i d_{ox}/\epsilon_{ox}[1-\exp(-\sigma_i jt/q)] \qquad (2.4)$$

where d_{ox} is the oxide thickness, t the injection time and j the current density. The various capture cross sections can be determined from the slope of the linear portions of the $\ln(d\Delta V_{fb}/dt)$ versus time curve. An alternative more accurate procedure, however, uses the Gauss-Newton method based on Jacobian matrices (81). This is much less sensitive to noise, finds always a least squares fit and is not limited with respect to the number of traps it can fit in one calculation. This method has been implemented in a fitting program written at ESAT[++].

2.3.3 Avalanche hole injection

With an n-type substrate avalanche breakdown results in injection of holes into the SiO_2. The mechanism is the same as that described for electrons in the previous section. One major problem exists with hole injection, however. This is the possibility of simultaneous injection of electrons from the gate so that the total current is due to both holes and electrons flowing through the sample. It has been shown by Aitken and Young (75) that the use of a sawtooth waveform is necessary to prevent electron injection. This ensures that whenever there is a high oxide field present there is also a displacement current in the silicon. Hence electrons need not be injected from the gate to satisfy current continuity requirements. Recent work has also shown that provided that the change in applied voltage needed to maintain a constant set current tracks the change in the flatband voltage, i.e. that no 'turnaround' effect is observed in the applied voltage, then negligible electron injection is occuring (82).

[++]Program written by M.Heyns.

2.3.4 Internal photoemission.

Using a transparent gate electrode, electrons excited by the absorption of u.v. light can be injected into the SiO_2 from either the gate or the substrate depending on the polarity of the applied bias (76,77). For an electron to be injected, it must first reach and then surmount the interfacial barrier. The probability of reaching the barrier depends on the likelihood of scattering, which in turn depends on the distance of the potential energy barrier maximum, x_{mo}, from the injecting interface (Fig.2.5). The probability of electrons surmounting the barrier depends both on the barrier height and on the energy with which they arrive at the barrier. Both the barrier height and position are affected by oxide charge and by the gate bias. Either the barrier height or its position can be made to dominate by using photon energies close to or considerably greater than the energy barrier, respectively. Oxide charge affects the barrier position much less than its height. However in the barrier height mode photocurrents are low and difficult to detect. The barrier position mode is therefore easier to use in practice.

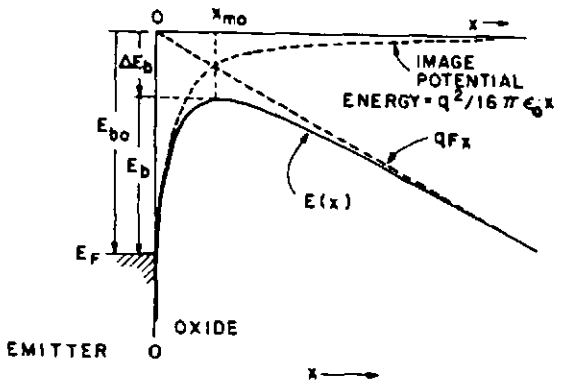


Figure 2.5 Effect of the image force potential on the potential energy barrier at the injecting interface (after 83).

Difficulties also arise in the analysis of the current characteristics in the barrier height mode since it is not certain whether tunneling through the top of the barrier should be taken into account (84).

Figure 2.5 shows an energy distance diagram of the $Si-SiO_2$ interface and the effect of the image force potential on the barrier. It can be shown that the position of the barrier maximum, x_{mo}, in the case of no oxide charge is given by:

$$x_{mo} = (qx_{mo}/16\pi\varepsilon_{ox}(V_g-\psi_{ms}))^{0.5} \qquad (2.5)$$

where ψ_{ms} is the metal-silicon work function difference (77). Thus, the barrier moves closer to the interface as V_g is increased.

Three cases are distinguishable when oxide charge is present. First if the charge is very close to the interface i.e. before the barrier maximum, the experimental photo I-V curve cannot be distinguished from that with no oxide charge present (Fig.2.6a). If the barrier maximum is within the charge distribution, this can be profiled. By varying the gate bias, the barrier maximum is swept through the oxide charge distribution and the resulting photo I-V curve is distorted compared with the case of no oxide charge (Fig.2.6b). Finally, if there is oxide charge only beyond the barrier maximum then the photo I-V curve is translated along the voltage axis remaining parallel to the curve in the case of no oxide charge (Fig.2.6c). The charge density, Q_o and the centroid, x_o can be extracted from the voltage shift measured first with the metal (ΔV_{gm}) and then with the silicon (ΔV_{gs}) as the injecting electrode (35), using the following equations:

$$\Delta V_{gs} = Q_o x_o/\varepsilon_{ox} \qquad (2.5)$$

$$\Delta V_{gm} = Q_o[(d_{ox} - x_o)/\varepsilon_{ox}] \qquad (2.6)$$

where d_{ox} is the oxide thickness and ε_{ox} the dielectric constant.

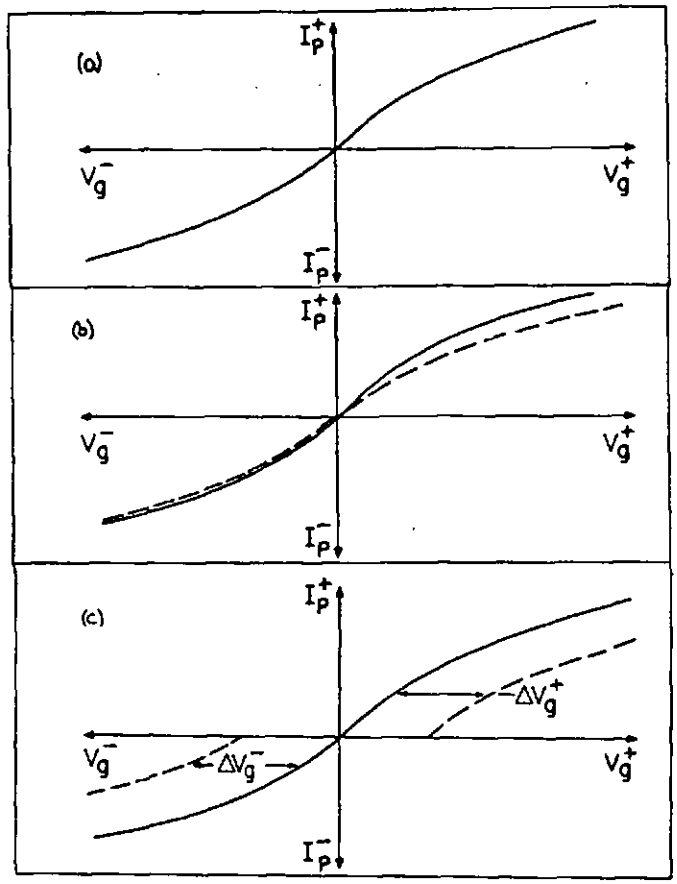


Figure 2.6 Shifts in the photo I-V curves for the case of (a) only interface charge, (b) charge at x_{mo}, (c) bulk charge. (After (85)).

The photo I-V technique is therefore a valuable means of examining charge distributions in the oxide. Currents may be kept low to avoid further charge trapping and at relatively low light intensities photodepopulation is also avoided.

2.3.5 Measurement systems.

Two automated systems have been developed in ESAT to study charge trapping effects in insulators. These are the avalanche injection and photo I-V systems. Both are connected to the same data acquisition system and HP 9835 controller. A switch allows the sample probe station to be easily connected to either system. A photograph of the measurement equipment is shown in Fig.2.7.

Figure 2.7 The avalanche and photo I-V measurement systems.

2.3.5.1 Avalanche injection system.

A diagram of this system is shown in Fig.2.8. Constant current d.c. stress may be carried out as well as avalanche injection of electrons or holes from the substrate. Either a d.c. voltage or a sawtooth waveform is applied, respectively and a feedback circuit regulates the voltage level or pulse amplitude to maintain constant current injection. At intervals during the stress the voltage is automatically measured at a set capacitance ratio, C/C_{ox}. This ratio can have any value from 0 to 1 but it is usually chosen as the value corresponding to either the flatband or midgap capacitances. Measurement of this voltage may take place every 10, 20, 50, 200 or 500 seconds. Six seconds are required to make the measurement regardless of the chosen time interval. The waveform used for avalanche injection is shown in Fig.2.9.

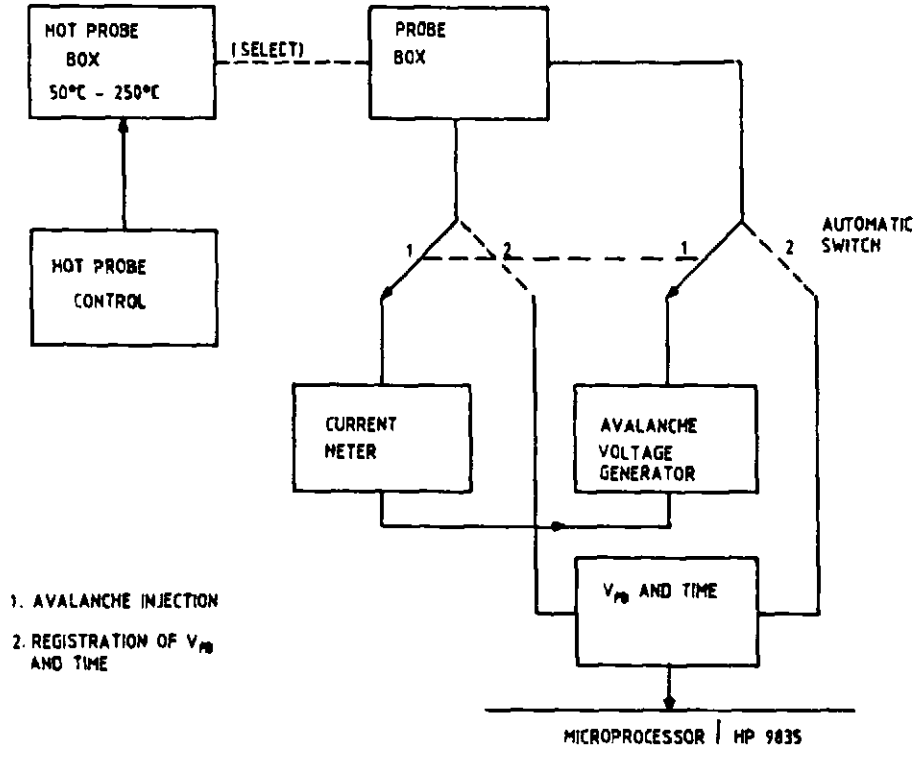


Figure 2.8 Block diagram of the avalanche injection system.

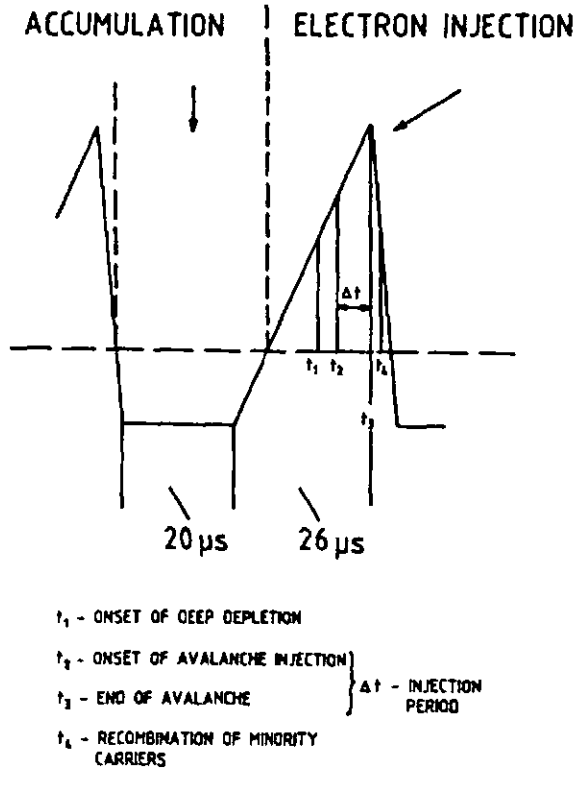


Figure 2.9 Sawtooth waveform used for avalanche injection.

Injection occurs only between times t_2 and t_3 so that 'packets' of current are injected. The electrometer, however, measures the average current through the device. This is compared with the set current and the amplitude of the pulse is adjusted accordingly. A sample-and-hold unit generates the ramp voltage for the C-V measurements, compares the measured capacitance value with the C/C_{ox} ratio chosen and determines the appropriate voltage. A timer measures the cumulative avalanche injection time and a driving pulse ensures that the correct measurement sequence is achieved.

2.3.5.2 Photo I-V system.

Figure 2.10 shows a schematic of this system. It consists of an optical bench with a xenon lamp, power supply and monochromator (H25 Jobin-Yvon), a high speed picoammeter (Keithley 417), a capacitance meter (Boonton 72BD), a precision lock-in amplifier (EG and G Brookdeal 9503D-SC), a microvoltmeter (Keithley 177) and an electrometer (Keithley 616).

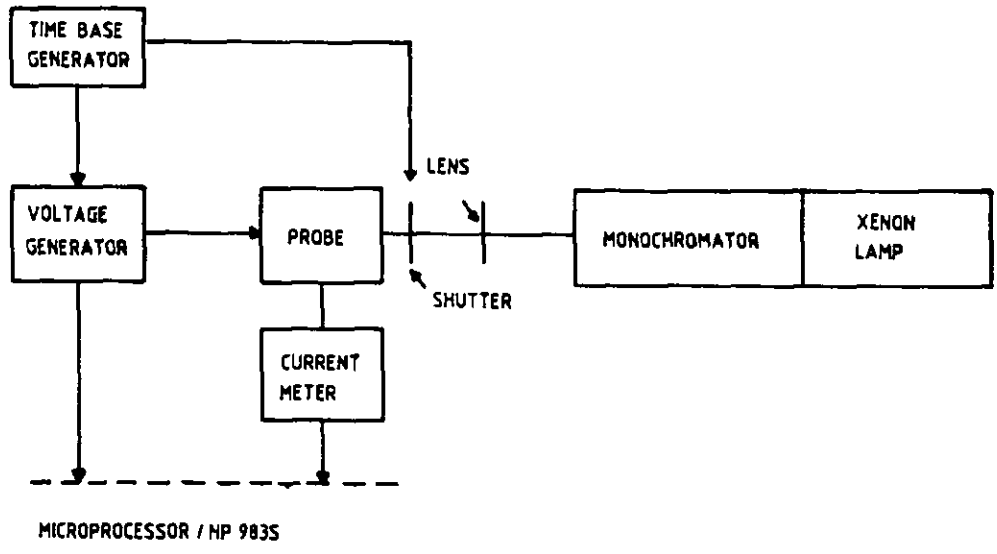
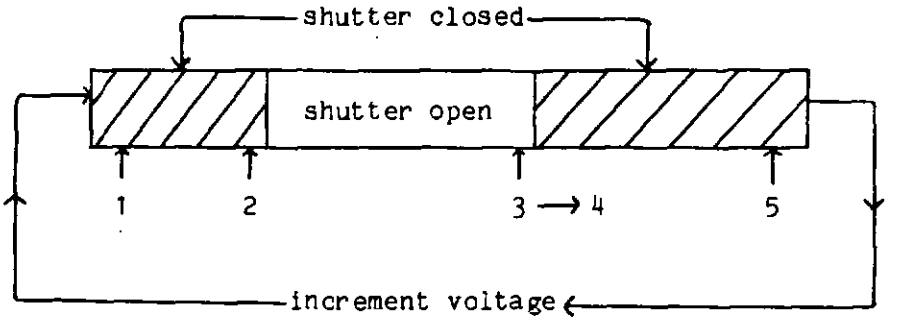


Figure 2.10 Block diagram of the photo I-V system.

The photocurrent through the MOS capacitor is measured for a specific value of the gate voltage. This voltage is applied in a series of steps; the magnitude as well as the duration of each voltage step (one measurement cycle) can be set from 5 seconds to 20 minutes. The voltage change between two sequential cycles is continuously variable from 0.1 to 11 V. During each cycle the sample is illuminated for a period which may be from 0.1-0.8 times the cycle time. Wavelengths of 2760 A (4.5 eV) and 2480 A (5eV) are normally used to cause injection from the metal and from the silicon respectively; i.e. the measurement is conducted in the barrier position mode. The sequence of measurements occurring in one cycle is illustrated in Fig.2.11 below.



1: At the start of the cycle the voltage is measured.
2: Before the shutter is opened the 'dark current' is measured.
3: Just before the shutter closes the 'light current' is measured.
4: 3 - 2, the light current minus the dark current is calculated.
5: At the end of the cycle the dark current is measured again.

Figure 2.11 Sequence of events in one cycle of a photo I-V measurement.

2.4 Capacitance-voltage (C-V) and charge pumping measurements.

High frequency (HF) and quasi-static (QS) C-V measurements and their analysis to calculate interface state densities are discussed extensively in ref. (11). These measurements can be made on both the

photo I-V and avalanche injection systems since both include a Boonton capacitance meter, a voltage ramp generator and a current meter. It is not necessary to change connections or sample holders. Figure 2.12 shows the principle of the two measurements. The quasi-static method is used to obtain a low frequency C-V curve. The displacement current generated as a result of the application of a slow voltage ramp is measured as a function of gate voltage. Since I = CdV/dT, the low frequency capacitance at the applied gate voltage can easily be calculated (86,87).

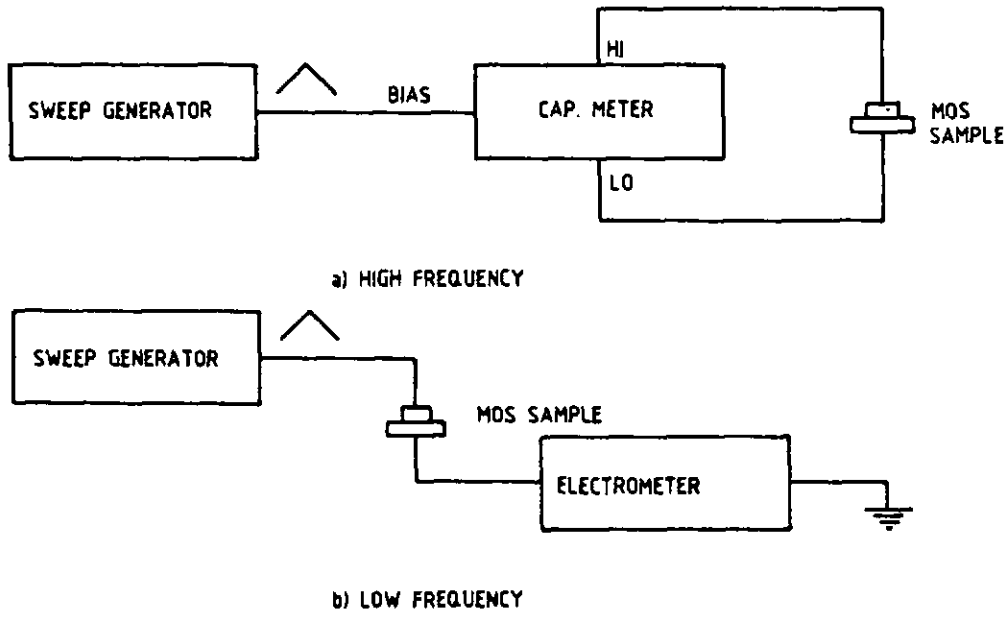


Figure 2.12 Principle of high and low frequency C-V measurements.

The time interval at which the capacitance or current is sampled can be varied from 0.3 to 20 seconds. Varying the ramp rate and the time interval determines the number of points taken in one voltage sweep. Data is taken automatically and stored on floppy disc for subsequent analysis. Programs are available to calculate the interface state density as a function of the surface potential or energy level within the silicon band gap using the methods outlined below in 2.4.3 and 2.4.4.

2.4.1 High frequency C-V measurements.

For an ideal MOS capacitor, the measured capacitance is due to the series combination of the oxide layer capacitance, C_{ox}, and the silicon depletion layer capacitance, C_s. The resulting value is a function of the gate voltage and measurement frequency. A normalised high frequency (1MHz) C-V curve calculated for an ideal p-type sample with an acceptor concentration of 2×10^{16} /cm^3 and an oxide thickness of 100 nm is shown in Fig.2.13. The electrode work function difference has not been taken into account.

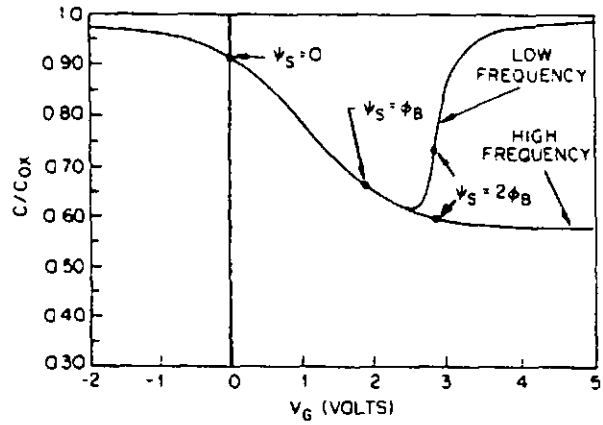


Figure 2.13 Calculated high and low frequency MOS C-V curves for a p-type sample (taken from (11)).

Three regions can be distinguished: accumulation, depletion and inversion. When the majority carriers are pulled towards the Si-SiO_2 interface by the applied gate bias the device is said to be in accumulation ($\psi_s<0$ in Fig.2.13). The capacitance then tends towards its maximum value, C_{ox}, where $C_{ox} = \varepsilon_{ox}A/d_{ox}$. ε_{ox} is the static dielectric constant of SiO_2, A is the area of the capacitor and d_{ox}, the oxide thickness. When the gate voltage is such that the majority carriers are repelled from the interface into the bulk of the silicon, ionised dopant atoms are left behind and a depletion layer is said to form ($2\phi_b>\psi_s>0$). The depth of this depletion

region and hence its capacitance varies with the gate voltage. At the onset of inversion, minority carriers begin to accumulate at the interface forming an inversion layer ($\psi_s > 2\phi_b$). The depletion layer approaches a maximum beyond which a small change in gate voltage produces a large increase in the number of minority carriers at the interface. However, under high frequency conditions, minority carrier generation and recombination is not fast enough to respond to the applied a.c. signal. The inversion layer therefore does not contribute to the measured high frequency capacitance, C_{hf}. Once the depletion layer is at a maximum, C_{hf} levels off at its minimum value. At any gate voltage:

$$C_{hf} = C_s C_{ox}/(C_s + C_{ox}) \qquad (2.7)$$

The flatband voltage of an ideal device is equal to the work function difference, ϕ_{ms}, between the gate material and the substrate. When bulk or interface charges are present it occurs at a voltage which compensates these charges. Hence:

$$V_{fb} = - Q_f/C_{ox} - (Q_{ot} + Q_{it})/C_{ox} + \phi_{ms} \qquad (2.8)$$

Charge in states whose occupation is independent of applied bias (Q_f and Q_{ot}) will cause a translation of the C-V curve along the voltage axis. Charge in interface states (Q_{it}) causes displacement and also distortion of the C-V curve along the voltage axis. Although the interface states make no direct contribution to C_{hf} since they cannot follow the a.c. signal, they are able to change their charge state in equilibrium with the d.c. bias. This causes a change in the surface band bending which in turn affects C_s resulting in 'stretch-out' of the C-V curve along the voltage axis. The flatband voltage shift relative to an oxide with no charge is given by:

$$\Delta V_{fb} = qN_{eff}d_{ox}x_o/(\varepsilon_{ox}d_{ox}) \qquad (2.9)$$

where N_{eff} is the effective charge density, x_o the charge centroid

and the other parameters are as previously defined. An unambiguous separation of the oxide trapped charge, Q_{ot} from the interface charge, Q_{it}, is not possible because the sign of the charge occupying the interface traps is indeterminate (88,89). The midgap voltage, V_{mg}, (i.e. the voltage when the Fermi level is at midgap) has been used to try and separate Q_{ot} and Q_{it} on the assumption that the net interface charge is zero at this point and hence any shift in V_{mg} is a measure of Q_{ot} alone (47-50). This is discussed further in 4B.1.2. V_{mg} is found experimentally as that voltage at which the depletion layer capacitance is given by the following equation:

$$C_s = [(q^2\varepsilon_s N_d/2kT)\ln[N_d/n_i]]^{0.5} \qquad (2.10)$$

where N_d is the doping density, n_i the intrinsic carrier concentration in silicon and ε_s the dielectric constant of silicon.

2.4.2 Low frequency and quasi-static (QS) C-V measurements.

Under quasi-static conditions or at low frequencies where the interface states are in equilibrium with the applied a.c. bias there is a direct contribution from these states to the capacitance, C_{lf}. This results in an additional capacitance, C_{it}, in parallel with C_s. Hence:

$$1/C_{lf} = 1/C_{ox} + 1/(C_s + C_{it}) \qquad (2.11)$$

A low frequency curve calculated for an ideal capacitor is also shown in Fig.2.13. Throughout the accumulation and depletion regions and until the onset of inversion, the capacitance is similar to the high frequency case. In inversion, because the minority carriers at the interface are now able to maintain equilibrium with the applied a.c. signal the measured capacitance rises again, once more approaching its maximum value of C_{ox}.

2.4.3 D_{it} distribution from QS and ideal C-V curves.

Rearranging equation (2.11) above we can obtain an expression for C_{it}:

$$C_{it} = [1/C_{lf} - 1/C_{ox}]^{-1} - C_s \qquad (2.12)$$

If C_{ox} and $C_{lf}(V_g)$ are measured and C_s calculated as a function of ψ_s, the variation of ψ_s with V_g is all that is required to obtain $C_{it}(\psi_s)$. Berglund (90) showed that:

$$\psi_s = \psi_{so} + \int_{V_{go}}^{V_g} [1-C_{lf}(V_g)/C_{ox}].\, dV_g \qquad (2.13)$$

V_{go} is an initial gate bias corresponding to ψ_{so}. The integral is determined experimentally from the low frequency C-V curve. Since ψ_{so} is not known, V_{go} is chosen to minimise the error in ψ_{so}, i.e. it is taken in accumulation or inversion where band bending is only a weak function of gate bias.

2.4.4 D_{it} distribution from QS and HF C-V curves

This method is due to Castagne and Vapaille (91). It has the advantage that C_s is determined experimentally from the HF C-V curve at the same value of the gate bias at which $(C_s + C_{it})$ is measured in the low frequency curve. Hence C_s automatically corresponds to the correct band bending. Also no theoretical calculation is necessary and no doping profile measurement is required. From the high frequency curve:

$$C_s = (1/C_{hf} - 1/C_{ox})^{-1} \qquad (2.14)$$

Substituting in the equation for the low frequency curve gives:

$$C_{it} = (1/C_{lf} - 1/C_{ox})^{-1} - (1/C_{hf} - 1/C_{ox})^{-1} \qquad (2.15)$$

Thus, C_{it} is readily obtained as a function of gate bias. ψ_s as a function of V_g can again be obtained from the Berglund integration. To obtain D_{it} over the silicon bandgap, the trap level E_t can be found from:

$$(E_c - E_t)/q \text{ [or } (E_t - E_v)/q] = E_g/2q + \psi_s - \phi_b \qquad (2.16)$$

2.4.5 BTS measurement system.

BTS measurements were carried out using a Model 410 EG and G PAR C-V plotter, a Temptronic TP36B thermochuck system with Karl Süss probes in a light tight box and a Houston 2000 x-y recorder (Fig.2.14). Temperatures of up to 305°C can be maintained with an accuracy of ±6°C and a stability of ±1°C. Heating is by a.c. heaters and cooling by an ethylene glycol/water solution. The temperature is sensed by a thermistor in thermal contact with the chuck surface. The stress voltage is continuously variable from zero to ±100V.

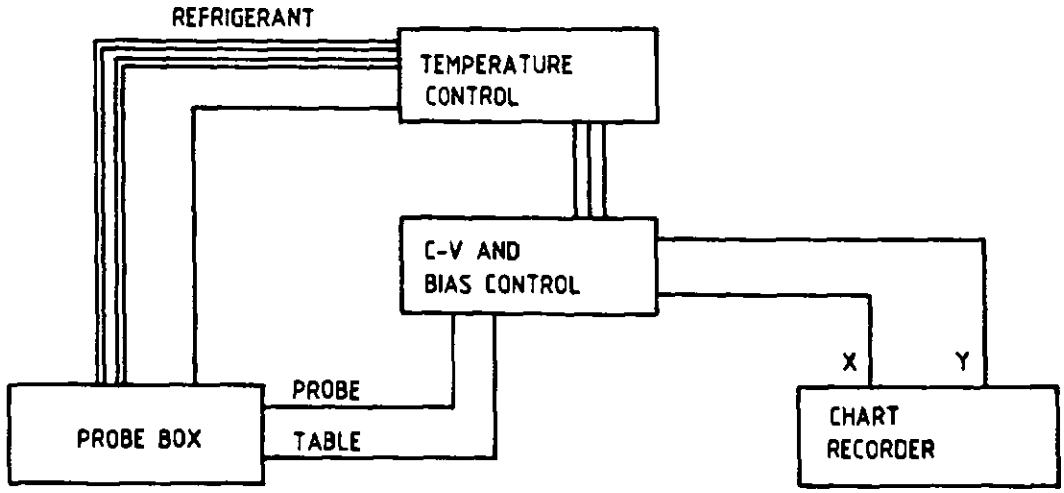


Figure 2.14 Block diagram of the BTS measurement system.

It is applied during heating and cooling as well as while the device is at the set temperature. Capacitance ranges of 0.5-2000 pF are available with an accuracy of ±2% of the measured value. A 1 MHz oscillator provides the 15 mV rms a.c. signal to the device under test. A typical measurement would consist of taking a room temperature C-V curve, heating to the set temperature (150-250°C) under stress (usually 2-4 MV/cm), maintaining this temperature for the chosen stress time, cooling to room temperature, switching off the bias and measuring a second C-V curve. The shift in V_{mg} between the two curves is then measured.

2.4.6 Charge pumping measurements.

All the methods to determine interface state densities, which have been discussed so far, require the use of MOS capacitors. The charge pumping technique, however, allows the measurement of interface state densities on MOS transistors (92). The circuit used to conduct a charge pumping measurement is illustrated in Fig.2.15 for the case of an n-channel MOST.

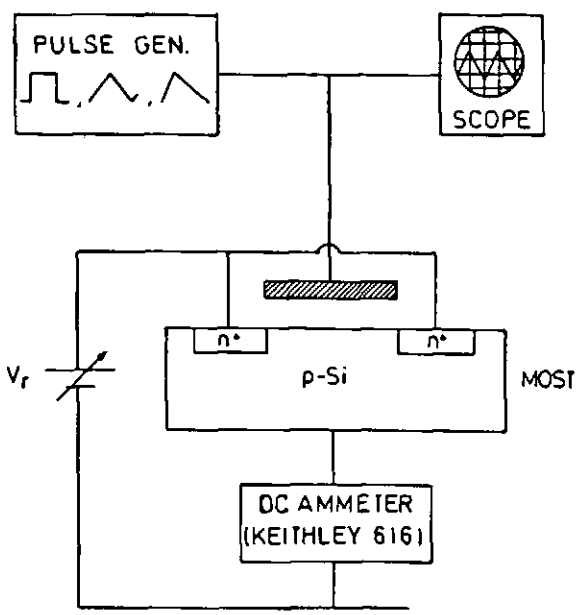


Figure 2.15 Principle of a charge pumping measurement. (After (93).)

In these measurements the pulse generator used was an HP 8010A and the oscilloscope a Philips PM 3217. The source and drain are connected together and reverse biased (0.5-1.0 V). A pulse waveform is then applied to the gate. When the transistor is pulsed into inversion the surface becomes deeply depleted. Electrons flow into the channel from the source and drain and some of them are captured by interface states. As the gate pulse drives the surface back into accumulation, the mobile charge drifts back to the source and drain under the influence of the reverse bias. Charge trapped in the interface states, however, will recombine with majority carriers from the substrate giving rise to a net flow of negative charge into the substrate given by:

$$Q_{it} = A_g q^2 \overline{D_{it}} \psi_s \qquad (2.17)$$

where $\overline{D_{it}}$ is the mean surface state density averaged over the energy levels swept by the Fermi level, A_g is the channel area and ψ_s is the total sweep of the surface potential expressed in eV. For repeated pulses of frequency f, the resulting substrate current is:

$$I_{cp} = fQ_{it} = fA_g q^2 \overline{D_{it}} \psi_s \qquad (2.18)$$

Measurement of this current therefore allows an estimate to be made of D_{it} over the energy range swept by the pulse.

A more accurate model for the charge pumping phenomenon was recently developed (93) and this technique now provides a simple quantitative method of measuring interface state densities on small gate area MOSFETs. Both an average value of the interface state density and the energy distribution of these states in the silicon bandgap can be obtained without knowing the dependence of the surface potential on the gate voltage. To obtain the energy distribution of the interface states the processes occurring on application of the gate pulse must be considered in more detail.

The waveform shown in Fig.2.16 is applied to the sample. It has a rise time, t_r and fall time, t_f, an amplitude ΔV_g and a period T. When the surface is in accumulation, the interface states below the quasi-Fermi level of the minority carriers are filled with electrons while those above it are empty. The states are thus in equilibrium with the energy bands. As the gate voltage and hence the surface potential increases, holes are emitted towards the valence band to maintain equilibrium and flow back into the substrate.

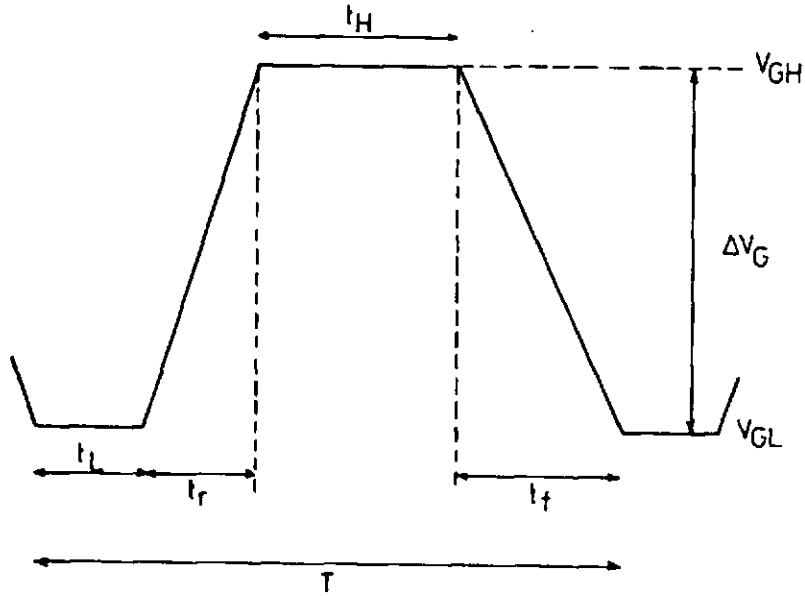


Figure 2.16 'Square' waveform used for the charge pumping measurements.

Initially the rate of emission of trapped charge to the substrate is able to keep trap occupation in dynamic equilibrium with the voltage sweep. As soon as the rate of change of the surface potential exceeds the emission rate, a non-steady-state regime is entered where trap emptying is completely controlled by the emission process (94). Because of the shape of the ψ_s versus V_g curve and the relatively high frequencies required to measure a significant substrate current, the transition point between steady-state and non-steady-state will always be very close to the V_{fb}. When the gate voltage is close to V_T, the trapping time constant becomes smaller and electrons will be trapped in interface states not yet emptied of holes. When V_g exceeds V_T, the remaining traps will be filled by electrons coming

from the source and drain. The channel is now back in equilibrium with the energy bands.

Similar mechanisms are operating as the gate is pulsed back. First electrons are emitted from the interface states and flow back to the source and drain in a steady-state regime. Then, in the depletion region non-steady-state emission of electrons to the conduction band followed by removal via the source and drain takes over. Finally when V_g equals V_{fb}, the trapping time constant of holes becomes important and these fill the remaining traps occupied by electrons.

According to the theory of carrier emission from interface states (94), the following expression can be derived for the charge pumping current in the non-steady-state regime using a 'square' waveform:

$$I_{cp} = 2q\overline{D_{it}}fA_g kT[\ln(v_{th} n_i \sqrt{\sigma_n \sigma_p}) + \ln(|V_{fb}-V_T|/|\Delta V_g|\sqrt{t_f . t_r})] \qquad (2.20)$$

where v_{th} is the thermal velocity of the carriers and σ_n and σ_p are the capture cross sections for holes and electrons respectively. By using varying rise and fall times it is possible to obtain the interface state distribution over the forbidden energy gap. When applying pulses with variable fall times and a constant rise time one scans the energy range in the upper half of the band gap and vice versa (93,95).

CHAPTER THREE:
BREAKDOWN OF SILICON DIOXIDE

3A: LITERATURE SURVEY

The stochastic nature of breakdown (96) means that there is no definite breakdown field or time-to-breakdown associated with a given device. At a constant applied field, even for identical devices, breakdowns are distributed in time and in a ramp test they are distributed in field. For identical devices, however, the probability of breakdown is constant. At defects this probability is increased so that breakdowns are distributed over lower fields in a ramp test and earlier times in a wearout test. As a result, both fast voltage ramp and wearout measurements show statistical distributions with more than one component. The low and medium field breakdowns in the first case and the early failures in the second are defect-related and due to weak spots in the oxide. The high field breakdowns or those occurring at longer times have often been assumed to be an intrinsic property of the oxide. However, it is by no means certain that this is actually the case. Over the past fifteen years there has been a substantial increase in the maximum value of the breakdown field, $E_{bd}(max)$, from 6 to 16MV/cm (18) and there is no reason to suppose that the limit has yet been reached. Part of this increase is due to a decrease in the oxide thickness used (5). Cleaner processing has also had a substantial influence though, perhaps accounting for 30 or 40% of the increase in $E_{bd}(max)$. Nevertheless, we shall continue to use the term 'intrinsic breakdown' as a convenient means of indicating both those breakdowns occurring at high field in a dielectric strength measurement and at long times to failure in a wearout measurement. Both intrinsic and defect-related breakdown have been reviewed by Solomon (97) and intrinsic breakdown mechanisms by Klein (98).

In the next section, the probable causes and the processing dependence of defect-related breakdown will be discussed. There follows a similar description of intrinsic breakdown and of the

physical models found in the literature for this phenomenon. Finally, the statistics of breakdown will be considered.

3A.1 Defect-related breakdown.

The most common technique for detecting oxide defects is electrical testing. Shorting or non-shorting tests may be employed. In the latter the upper electrode is thin enough to evaporate from the region above the defect when breakdown occurs. The capacitor is not destroyed by this 'self-healing' breakdown and many tests can be carried out on a single large device. However, this technique assumes that after a self-healing breakdown the measurement continues without any influence from the destruction of part of the insulator. Since this assumption is dubious, shorting tests using large matrices of small capacitors are generally preferred.

Various other techniques have also been employed to measure the defect density in SiO_2 (99). These include liquid crystal methods which can non-destructively detect regions of high current density (100,101), scanning internal photoemission (102,103) and scanning internal photovoltage (104) techniques. Defects at the Al-SiO_2 interface have also been inferred from deviation of the I-V characteristics from the expected behaviour as a result of barrier height inhomogeneities (105).

3A.1.1 Nature of defects.

Defect density in SiO_2 is affected by several factors, including substrate quality and wafer cleaning procedures (5,16,99), cleanliness of the oxidation furnace (5) and metallization conditions (16). The nature of the defects varies considerably. Particulate contaminants embedded in the oxide can cause a localized reduction in dielectric strength or when on the substrate surface they may interfere with the film growth causing thin spots, partial or complete pinholes or hillocks. Pinholes may also result from photolithographic processing

if the photoresist does not completely protect the film thus allowing etchants to penetrate through flaws. Stress may also be a cause of defects. This can develop in films during growth, deposition, pattern etching or heat treatments. Aluminium migration (16) and sodium contamination (106,107) have also been implicated in reduction of the breakdown field. Metallic impurities from the furnace heating elements and other sources have in particular been associated with medium field breakdowns (108,21). A back side diffusion of phosphorus was found to getter any metal ions and to substantially reduce this failure mode (108), while use of a double-walled oxidation tube lowered the level of metallic contaminants incorporated during film growth (21). High levels of oxygen precipitates present in some as-received wafers have also been shown to result in more medium field breakdowns (108). A pre-oxidation anneal in nitrogen helped to reduce these.

This diversity of types of defect is probably largely responsible for the inconsistency in the results reported for the effect of processing parameters on defect-related breakdown. For example, it has been variously reported that increasing oxide thickness causes a decrease (16), an increase (109) or that it has no effect (5) on the defect density. Recent work suggests that this may be explained by low field breakdowns decreasing with oxide thickness while medium field breakdowns increase (108). If all defect-related breakdowns are considered together, the thickness dependence will depend on which of the two modes dominates.

3A.1.2 Processing and materials dependence.

Table 3.1 shows a summary of the published results on the relationship of defect density to materials and fabrication conditions. Annealing conditions, oxidation temperature, addition of HCl to the oxidation ambient and a phosphosilicate glass (PSG) passivation layer were found to be the most significant parameters. A dramatic increase in defect density is observed following an extended

PMA (~1h, 500°C) with magnesium or aluminium gate electrodes (16) but not with gold (16) or polysilicon (5). Aluminium was found to react with the SiO_2 at these temperatures eventually causing shorting. A short PMA (5', 500°C) on the other hand gave a slight improvement in oxide integrity probably due to a reduction in interface state density (5). Similar results have been obtained for post-oxidation anneal conditions. Osburn and Ormond (5) found that a high temperature POA in nitrogen (15 min., 1000°C) increased the defect density by 50% in 1000 A films and by 400% in 200 A films. In general, the defect density would appear to increase with increasing anneal time (5,110). However, recent work has shown that a short POA (10') at the oxidation temperature (900-1000°C) does initially decrease the defect density (111).

Parameter (increased)			Defect Density
Substrate	:	doping	↑ (5)
Oxidation	:	temperature	↑ (15,16,111); ↓ (112)
	:	ambient water conc.	= (5)
	:	ambient HCl conc.	↓ (5,113)
	:	oxide thickness	= (5); ↑ (109); ↓ (16)
	:	PSG layer	↓ (5,16)
Anneals	:	POA time (N_2)	↑ (5,110)
	:	PMA time/temp.	↑ Al,Mg (16); = poly,Au (5,16)
Electrode	:	(Al,Mg,Cr,Mo,Au,Pt)	= (5)

Table 3.1 Summary of the reported results on the processing dependence of the defect density in SiO_2 layers.

The balance of evidence suggests that higher oxidation temperature favours an increase in the defect density (5,16,111) but the amount of water in the ambient does not seem to be important (5). Higher oxidation temperatures appear to have a more detrimental effect the lower the oxide thickness and to have relatively little effect on films greater than 1000 A (5).

Chlorine treatments are thought to remove or counteract the effect of metallic impurities such as sodium thereby reducing defect levels (113). HCl, both in the oxidation ambient and as a furnace pre-clean treatment reduces defects substantially (5) and increases defect-related wearout time (13). A PSG layer, if it is thicker than about 60 A, also reduces the defect level (5,16).

High substrate doping appears to cause a slight increase in defect density especially for thinner oxides. A 10-20% decrease in the usual breakdown strength was reported for a 200 A oxide in the extreme case of 10^{20} dopant atoms /cm^3.

It is interesting to note that there is some correlation between the effect of oxidation parameters on roughness of the $Si-SiO_2$ interface (114) and that on defect-related breakdown. A low oxidation rate (low temperature) favours a smoother interface and also reduces the defect density. Similarly a POA in N_2 reduces interface roughness and provided it is not too long it also improves the defect density. A PMA has little effect on the $Si-SiO_2$ interface and for polysilicon gate devices there is also no effect on the defect density. The increase in defect-related breakdowns for aluminium gate capacitors is almost certainly due to defect creation at the metal-oxide interface.

3A.1.3. Sodium-related breakdown.

Sodium ions are a common impurity in MOS devices unless extreme care is taken to ensure clean procesing conditions and there have been many studies of the effects of the mobile Na^+ on breakdown in SiO_2 (103,106,107,115-118). There is apparently no effect on the breakdown distribution at negative gate polarity (116,117). Time dependent breakdown at positive gate polarity due to Na^+ was first observed by Worthing (118) although he did not connect it with sodium. He fitted his results to an empirical relationship known as Peek's law which is of the form:

$$E_{bd} = E_o + at^{-0.25} \qquad (3.1)$$

By purposely contaminating samples with NaCl, this breakdown mode was attributed to sodium (106,116). As the Na^+ concentration increased, breakdown occurred at lower and lower fields, approaching 1.5 MV/cm for 10^{15} /cm^2 of sodium ions. On applying a positive gate voltage of 4 MV/cm a transient current peak was seen in sodium contaminated samples. Most breakdowns were found to occur at times close to the maximum of this peak. The premature breakdown was attributed to field enhancement and hence increased current injection caused by the sodium ions traversing the insulator. Once the ions are within tunneling distance of the Si-SiO_2 interface, they cease to affect significantly the current and therefore have less effect on breakdown. However, longer time breakdowns have been attributed to clustering of sodium at the interface causing localized barrier lowering. Williams and Woods (102) and DiStefano (103,107) used scanning internal photoemission to correlate regions of barrier lowering with sodium concentration and with breakdown.

3A.1.4 Area dependence.

Boltzmann statistics are usually used to describe the distribution of defects in the oxide. This assumes a random distribution of independent defects. To obtain the area dependence of the breakdown probability due to these defects a capacitor of area A may be considered as n smaller capacitors of area a connected in parallel. The probability of breakdown occurring in one of these regions is then equal to the probability of finding a defect in that area i.e. $p = Da$ where D is the defect density. By definition $p \leq 1$ and for small areas $p << 1$, therfore Da is also $<< 1$. The probability of the large capacitor surviving a given applied field is expressed in terms of the yield, Y, which is given by:

$$Y = (1-p_1)(1-p_2)(1-p_3)(1-p_4).....(1-p_n) \qquad (3.2)$$

If there is an equal chance of failure of all the small areas, then:

$$Y = (1-p)^n \qquad (3.3) \qquad \text{Since } p = aD,$$

$$Y = (1-aD)^n \qquad (3.4)$$

As aD is very small and consequently n large all terms above the first in the series expansion may be ignored so that $Y = (1-naD)$. Since $a = A/n$, $Y = 1-AD$. For $AD<<1$ this approximates to:

$$Y = \exp(-AD) \qquad (3.5)$$

This is the most commonly used relationship to extract a value of the defect density from the number of low field failures in a fast voltage ramp breakdown measurement.

An alternative formula for the yield has been used by some authors (109,115,119):

$$Y = 1/(1 + AD) \qquad (3.6)$$

This is based on Bose-Einstein statistics where defects are assumed to be indistinguishable. However, this equation violates the principle of statistical independence for parallel connected capacitors and there is also no reason to suppose that defects are indeed indistinguishable.

3A.1.5 Field, time and temperature dependence.

As shown above, the number of failures occurring at or below a given field can be used to calculate a value for the defect density. However, the defect density is, of course, itself a function of field, time under stress and temperature although this is not always recognised. D in eqn. 3.5 is not actually a measure of the number of defects present but of the density of defects which break down under

test. (Shatzkes et al. (105) termed this the breakdown density.) The test conditions are therefore crucial to the value of D which is obtained. Elevation of the field or temperature in a wearout test shifts the breakdown distribution to earlier times (14,38,39). Similarly, under ramp test conditions decreased ramp rates result in increased wearout and hence lower measured breakdown fields and higher values of D. Practical use is often made of this dependence on the measurement parameters to screen out potential early failures before they come into operation (120,121). If we examine the failure rate (number of failures per unit time) during a wearout measurement, it usually takes the form of a 'bathtub' curve (Fig.3.1).

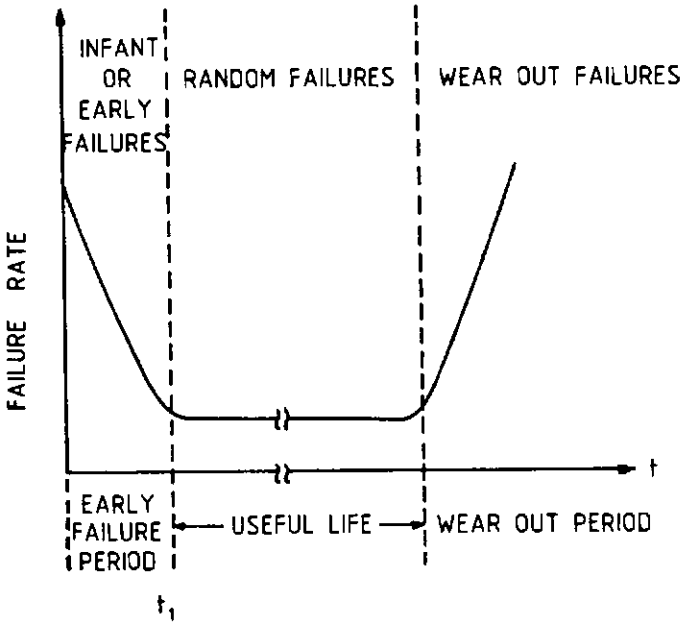


Figure 3.1 'Bathtub' curve of the failure rate as a function of time during a wearout measurement (after (120)).

The early failure region is characterised by a decreasing failure rate. This is followed by a low and constant failure rate before the bulk of the devices begin to fail. Screening or 'burn-in' consists of applying a stress equivalent to that at time t_1 in Fig.3.1 i.e. until all the weak devices have failed. Elevated temperature plus higher fields than would normally exist under operating conditions are generally used to speed up this process (14,38).

3A.2 Intrinsic Breakdown

3A.2.1 Processing and materials dependence.

Parameter (increased)			E_{bd}(max)	t_{bd}(max)
Substrate	:	doping	= (8); ↓ (5)	-
Oxidation	:	temperature	-	-
	:	ambient water conc.	= (5)	-
	:	ambient HCl conc.	↑ (113)	↑ (13)
	:	oxide thickness	↑ (16); ↓ (5,8)	↑ Al; = poly (13)
	:	PSG thickness	↑ (5,16)	↑ (13)
Anneals	:	POA time (N_2)	↓ (110)	-
	:	PMA (5'/N_2/500°C)	-	↑ (5)
Electrode material	:	(Al,Mg,Cr,Mo, Au,Pt,poly)	= (5,122)	↑ (13,19)

Table 3.2 Summary of reported results on the processing dependence of E_{bd}(max) and t_{bd}(max).

Table 3.2 summarises results from the literature on the effect of materials and processing parameters on the maximum breakdown field, E_{bd}(max) and on the maximum wearout time (t_{bd}(max)). The dependence of E_{bd}(max) on all parameters except the oxide thickness is slight. The average breakdown field, $\overline{E_{bd}}$, follows a similar trend as can be seen from Fig.3.2. This shows experimental results from the present work illustrating the increase in the $\overline{E_{bd}}$ with decreasing oxide thickness below 50 nm. Wolters (123) has measured E_{bd}(max) as a function of oxide thickness using a constant rate of ramping the field rather than the more usual constant voltage ramp rate to ensure that the contribution from wearout is the same for all thicknesses. The decrease in E_{bd}(max) with oxide thickness was still observed however.

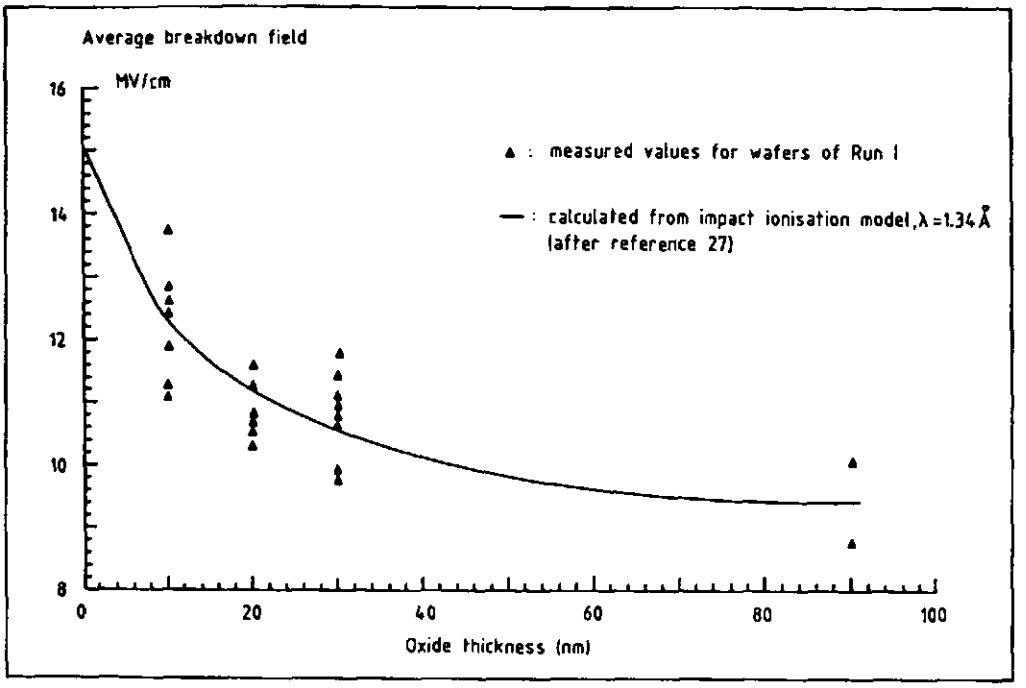


Figure 3.2 Variation in E_{bd} with oxide thickness.

Since E_{bd} increases with decreasing oxide thickness, Q_{bd}, the total injected charge prior to breakdown, must also increase. This has been confirmed by constant current wearout measurements where Q_{bd} is more easily measured (37). Apparently contradictory results have been reported from constant voltage wearout measurements. Here $t_{bd}(max)$ was found to remain constant with increasing oxide thickness for capacitors with polysilicon gate electrodes and to increase with aluminium electrodes (13,19). However, the relative amounts of charge trapping in these particular devices is not known. The variation in the net applied field, the injection current and hence Q_{bd} can therefore not be estimated.

Considering that the current is limited by F-N tunneling, there is remarkably little dependence of the maximum breakdown field on the electrode material. On going from gold to magnesium, for example, the

current varies by several orders of magnitude for the same (negative) value of the gate bias. There is negligible change in $E_{bd}(max)$, however (16). Ormond measured no significant difference at all between the breakdown distributions of polysilicon and aluminium gate devices made on the same wafer (122). On the other hand much longer times to breakdown were found with polysilicon rather than aluminium electrodes especially with the gate biased positively (19). From these results wearout would appear to depend primarily on the anode material. The trend is again not what might be expected from barrier height considerations since $E_{Al} > E_{Si}$.

Parameters which reduce the impurity level such as use of a PSG passivation layer or HCl in the oxidation ambient to getter metal ions have similiarly beneficial effects on $E_{bd}(max)$ (16) and $t_{bd}(max)$ (18). This again suggests 'intrinsic' breakdown is not being observed since $E_{bd}(max)$ is being influenced by removal of contaminants. As with defect-related breakdown some correlation can be found between surface roughness and $E_{bd}(max)$. A low oxidation temperature and a short POA improves the intrinsic breakdown and gives a smoother interface (114).

3A.2.2 Dependence on measurement parameters.

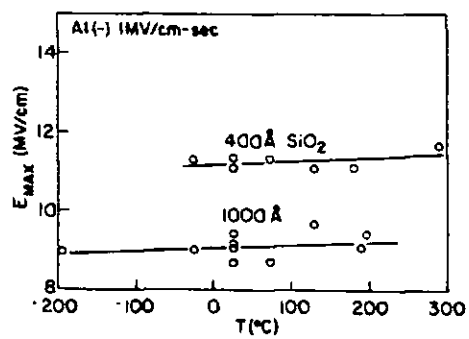


Figure 3.3 Temperature dependence of $E_{bd}(max)$ (taken from (5)).

The temperature dependence of $E_{bd}(max)$ is slight, but if anything it appears to increase with temperature (Fig.3.3). Harari found the dielectric strength did not change with increasing temperature but the time-to-breakdown under constant current injection conditions was shortened considerably (8). Constant voltage wearout measurements show a time to breakdown exponentially dependent on the temperature (14,38). An activation energy for wearout can be calculated from these measurements. This has recently been shown to be itself field dependent (124) thus accounting for the disparity of values previously reported in the literature (14,38,125,126).

There is a quasi-exponential decrease in t_{bd} with applied field (14,23,127). The field acceleration factor often used to quantify this relationship is also temperature dependent (124). Klein's measurements show that the $\ln(t_{bd})$ versus E curve is very steep for SiO_2 (127), which led many workers to overlook the time dependence of intrinsic breakdown i.e the effect of ramp rate on breakdown field distributions. The distribution shifts to lower fields as dV/dt is reduced.

3A.2.3 Physical models for breakdown.

A number of possible models for breakdown can be eliminated immediately because they are not compatible with experimental observations. For example, the relative temperature independence of the conduction mechanism (12,24) precludes thermal runaway of the whole sample (128). Ionic effects also cannot explain the time and temperature dependencies (129). Other mechanisms such as the electrochemical models of Budenstein (31,130,131) and Wolters (32,33,123) can certainly not be discounted but as yet they have relatively little experimental support. Other contenders, which are able to explain some observations are the electron trapping model of Harari (8), resonant tunneling (132) and hole injection from the anode leading to current instabilities (133). Probably the most succesful model to date in explaining the experimental observations is the

impact ionization model (6,134) However, there are problems with this also for very thin oxides. We shall consider a few of the most important models and discuss their successes and shortcomings.

3A.2.3.1 Impact ionization model.

An impact ionization model for SiO_2 was proposed by O'Dwyer (134). It was modified by DiStefano and Shatzkes (6) to include recombination and non-local ionization rates thereby producing a model which could explain the oxide thickness dependence and electrode material independence of the breakdown field. It consists of the following assumptions:

1. Electrons are injected into the SiO_2 conduction band from the cathode by F-N tunneling and are accelerated by the field.
2. A small fraction (~1 in 10^9) of the hot electrons generate electron-hole pairs by impact ionization of the lattice atoms. The electrons have a high mobility (135) and are rapidly swept out of the oxide while the relatively immobile holes (136) remain behind enhancing the cathode field and hence electron injection.
3. Drift and recombination of the holes opposes this feedback mechanism, recombination giving the electrode material independence of the breakdown field.

Forward scattering of electrons from optical phonons was considered and the electron-phonon mean free path was adjusted to give the limiting breakdown voltage of 92.5 V for a 100 nm oxide. At this thickness E_{bd} is approximately equal to $E_g/41q\lambda$. For thin films there is a strong thickness dependence, giving good agreement with experiment (6). The non-local ionization rate has two consequences. Firstly, it predicts a negative differential resistance range, which is observed experimentally and secondly no ionization is possible for a distance $x = (E_g + E_{bo})/E_{bd}q$ from the cathode. (E_{bo} is the

barrier height.) The increase of E_{bd} with temperature can be qualitatively explained by a decrease in either ionization or hole trapping.

A simplified version of this model expected to be valid for oxides >100 nm was treated by Klein and Solomon (29). They obtained analytical expressions for the threshold field for current runaway. Times to runaway at fields above threshold were calculated and found to be inversely proportional to the current and ionization rate and to decrease quasi-exponentially with the field. Thus, although the threshold field was electrode independent, time to runaway was exponentially dependent on the electrode work function. This is also at least qualitatively compatible with observations.

Objections to the impact ionization model have been raised on the grounds that at the measured breakdown fields negligible ionization can in fact occur in such a wide bandgap insulator (129,137). The predicted rate of ionization depends on some rather inaccurately known parameters such as the electron scattering and the precise SiO_2 band structure. Klein demonstrated that for wide band gap insulators the energy required for electron-hole pair creation is roughly 3 x E_g or in SiO_2, 27 eV (138). Calculations based on plasmon theory (139) yield much lower values of the order of 17 eV (140). Whichever value is correct, electrons in the hot tail of the energy distribution have been shown to attain these energies at fields below the breakdown field (18). However, no correlation was found between the samples with the hottest energy tails and earlier breakdown. This suggests that although the electrons reach energies sufficient to cause impact ionisation this is not the ultimate cause of breakdown. Probably the biggest argument against this model is its inapplicability to thin oxides where breakdown occurs at voltages below the minimum required for impact ionization i.e. $(E_g + E_{bo})/q$ or about 13 V in SiO_2.

3A.2.3.2 Anode hole injection model.

Since impact ionisation cannot be the cause of positive charge generation and breakdown in oxides below ~100 A, doubt is cast on the role it plays in thicker layers. However, positive charge build-up undoubtedly accompanies electron injection under high field stress. It has been shown that this positive charge is due to the trapping of holes in oxide hole traps already present prior to stressing (141). This charge may precipitate breakdown, either in a manner similar to that outlined above in the impact ionisation model or by damage creation accompanying the charge trapping/detrapping process. Hole injection from the anode has been suggested as a means by which these holes might enter the oxide (20,142,143,144,164). No well-developed breakdown model has yet been formulated based on this process but there is evidence that such hole injection does occur. Fischetti has recently shown theoretically that there is a significant probability of surface plasmon mediated hot hole injection for anode fields >5 MV/cm (144). The predictions of this theory are also consistent with the experimental data. In particular, the model is supported by the observation that the rate of positive charge generation (hole trapping) depends on the anode field (164). The creation of surface plasmons by hot electrons leaving the insulator has also been shown experimentally (145).

3A.2.3.3 Electron trapping model.

Harari (8,30) has proposed a model in which Si-O bond breaking causes electron traps to be generated in the oxide under high field stress conditions. Population of these traps results in large local fields in the oxide leading to further bond breaking and hence breakdown. From measurements at constant current, he monitored both the time to breakdown (t_{bd}) and the change in applied voltage at breakdown (ΔV_{bd}). The following observations were made:

1) ΔV_{bd} changed linearly with t_{bd} but the gradient of this line was less for the 44 Å oxide than for the thicker oxides. Harari suggests that electrons giving rise to ΔV_{bd} are trapped predominantly near the injecting electrode and only the 44 Å oxide allows tunneling directly out of the oxide giving a smaller net ΔV_{bd}.

2) Different duty cycles made no difference to the net t_{bd} indicating that the effects of stressing were cumulative.

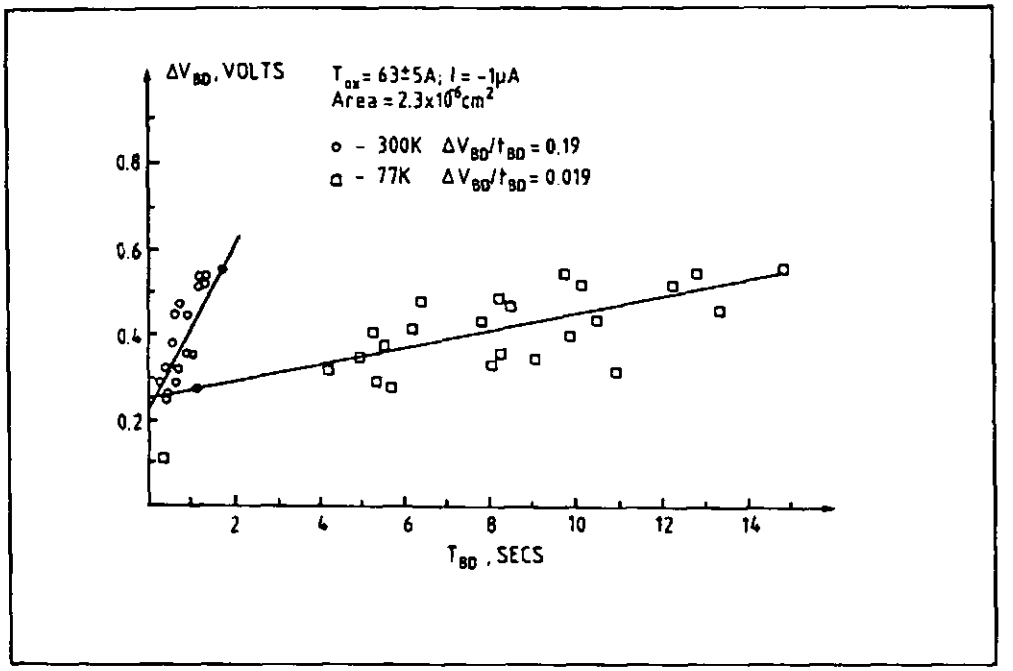


Figure 3.4 ΔV_{fb} versus t_{bd} at 77 K and 300 K (from ref.(8)).

3) Breakdown occurred ten times sooner at room temperature than at 77 K. ΔV_{bd} was very similar however. Hence a plot of ΔV_{bd} versus t_{bd} gave a line for which the slope was ten times greater at room temperature than at 77K (Fig.3.4). This implies that under very similar oxide stress conditions with identical current density, generation of traps and trapping of electrons is much faster at room temperature than at 77 K.

4) ΔV_{bd} did not saturate exponentially with time as it would if the traps had all existed in the pre-stressed oxide and were only being filled rather than created. The linearity of ΔV_{bd} with time implies rather a uniform rate of creation until breakdown occurs. Below 100 Å the electron trap density at breakdown was also the same for each of the films. The linearity of ΔV_{bd} as a function of oxide thickness (Fig.3.5) suggests that initiation of breakdown is localised at the $Si-SiO_2$ interface, while the saturation value reached at 100 Å implies the existence of a transition region at that thickness. The electron traps created are deep stable defects since under the high field conditions any electrons trapped within ~4 eV of the oxide conduction band (or within about ~30 Å of the anode interface) would be within direct tunneling distance.

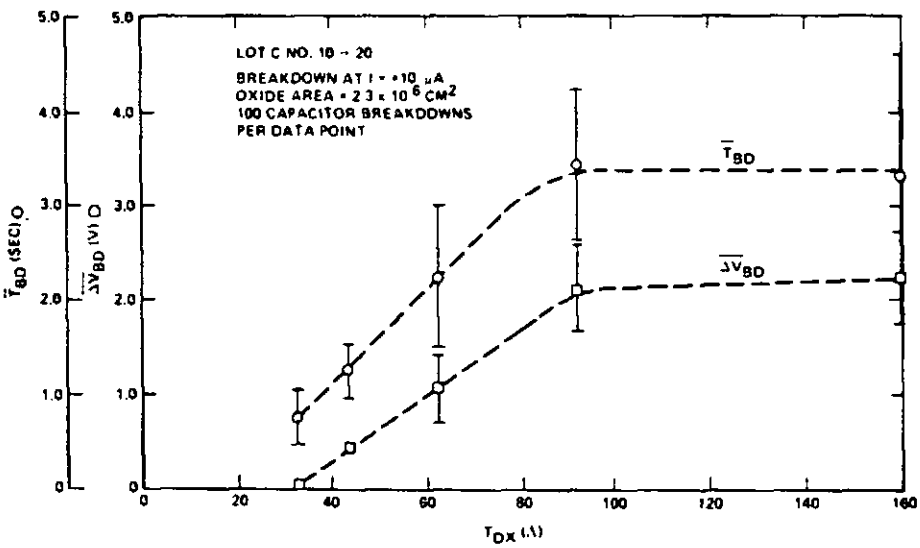


Figure 3.5 ΔV_{bd} and t_{bd} as a function of oxide thickness (from (8)).

Hole mobility in SiO_2 has a strong temperature dependence and if breakdown were due to impact ionization it should occur sooner at 77 K than at room temperature. However, the opposite result was found. ΔV_{bd} increases more slowly at lower temperatures and this may be due to compensation from trapped holes. Injected electrons undergo fewer interactions with phonons at lower temperatures, though and there is

consequently a slower rate for defect formation. The time-to-breakdown is also larger by almost the same factor.

It has subsequently been shown that trap generation precedes breakdown of highly stressed thermal SiO_2 (146). (These traps are located primarily at the non-injecting interface however.) Certainly in the thinnest films studied by Harari where breakdown occurs below 8V no impact ionization can take place. The very high breakdown fields of ~3 x 10^7 V/cm may well reflect a fundamental limit on bond strengths in the system. Ferry calculated that hot electron runaway should typically occur at this magnitude of the field (137). Harari's measurements were however made on exceptionally thin oxides (<170 A) using very large currents and small capacitors. It is possible that a change in breakdown mechanism takes place on going to thicker oxides, lower currents and hence lower fields. The transition at 100 A for t_{bd} as a function of thickness found by Harari himself is perhaps indicative of this.

3A.2.3.4 Resonant tunneling model.

A self-accelerating mechanism for current runaway by resonant tunneling into defect-related states has been proposed by Ricco et al (132). Trap assisted tunneling has been treated by other authors without taking into account its resonant nature (147). Resonant tunneling has already been demonstrated in double barrier structures (148). The only prerequisite for it to occur in normal MOS structures is the presence of localised states in the insulator energy gap, a condition which is commonly met. A rapid local increase of current occurs at the resonance condition and the accompanying thermal effects can lead to irreversible failure.

Defect states can be treated as wells in the potential energy barrier (Fig.3.6a). They are assumed to be strongly localized (i.e. both barriers have a transmission coefficient <<1) so that they behave as eigenstates of their energy wells.

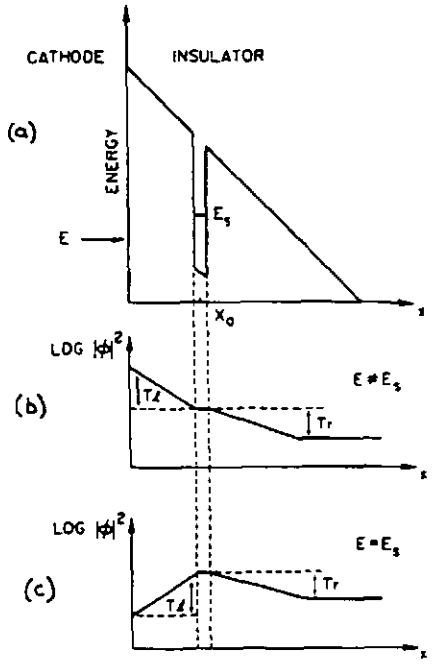


Figure 3.6 a) MOS potential energy diagram showing a defect-related eigenstate of energy E_s.

b) wavefunction of an electron incident on the barrier at an energy, $E \neq E_s$.

c) wavefunction of an electron incident on the barrier at an energy, $E = E_s$, (from (132)).

In the ideal case of a single eigenstate, E_s and a flux of noncharged, monoenergetic carriers of energy E incident on the cathode, if E does not match the energy of the eigenstate, the only effect is a small reduction of the barrier area (Fig.3.6b). The transmission coefficient of the total barrier, T_g, is equal to the product of the coefficients for the left and right barriers, $T_l.T_r$ ($= T_{min} \times T_{max}$). If the carrier energy matches that of the eigenstate, resonant tunneling occurs. Instead of decreasing exponentially within the classically forbidden area, the wave function is now peaked at the eigenstate (Fig.3.6c) and there is an accumulation of probability density in the well. T_g then equals T_{min}/T_{max}. For T_g to be of the order of one therefore it is only necessary for T_l and T_r to be almost equal. Since the concentration of defects is expected to be strongly peaked at the interface, at low field $T_l >> T_r$ and the eigenstates simply behave as fast traps for the tunneling carriers. As the field increases T_l

approaches T_r. At $T_g = 1$, the barrier opposing carrier transport no longer does so and there is essentially a short between the electrodes and breakdown. The field required for this is about 10 MV/cm in good agreement with the observed breakdown field for 1000 A oxides.

Charge trapping at defects increases the potential energy and hence E_s. It makes both T_r and T_1 larger and for defects near to the cathode more nearly equal. This increases the tunneling rate giving rise to a self-accelerating mechanism. Time to breakdown is expected to decrease with tunneling rate (and hence current) in good agreement with experiments. The model suggests that breakdown should be accelerated by increasing temperature, as is observed experimentally (6), because as the carrier energy distribution spreads the probability of finding the right eigenstate increases. (Only a few states will be in a position to fulfil the condition $T_1 = T_r$. The majority simply contribute to electron trapping.) For any single eigenstate the current should decrease slightly with increasing temperature as the concentration of available carriers at a given energy decreases. This could account for the observed slight increase in $E_{bd}(max)$ with increasing temperature.

3A.2.3.5 Gas discharge model.

Budenstein and co-workers (31,130,131) developed a general model for electrical breakdown in insulators based on the formation of a gaseous conducting channel. This has subsequently been elaborated upon by Wolters (32,33,123). The principle of the model can be summarized in five steps (130).

1. A critical charge density must be attained in a local region of the insulator. The source of this may be charge injection from an electrode, impact ionization or some other means.
2. Bond breaking is assumed to result from the effect of the charge density. The products include excited atoms and ions of several eV in energy as well as free electrons.

3. A chain reaction is set up in that freed electrons can cause other bonds to be broken in the solid. These in turn release atoms, ions and more electrons forming a gaseous plasma. Electrons striking the wall of the created cavity cause further dissociation. Gaseous collisions and photoabsorption may also produce additional electrons.
4. A gaseous channel is established by electrons striking the wall of the cavity and eroding it. The channels normally grow from the anode end (31) forming 'trees' as they run into each other.
5. Completion of a channel between the electrodes allows a conducting pathway to form. A local high current flows and heating causes the channel to enlarge further increasing conductivity until breakdown occurs.

There is some experimental evidence in favour of this model of breakdown for SiO_2. For example, time resolved spectroscopy of the light emitted during breakdown of Al-SiO-Al capacitors showed lines characteristic of single atoms and ions from the dielectric appearing within 20 ns of the onset of breakdown (131). Treeing is well documented for cable insulation materials where the trees begin at asperities, voids or imbedded particles all of which serve as local field enhancement sites. It has been observed directly in PMMA (31).

3A.3 Breakdown statistics.

Several statistical models have been used to describe the field and time distributions of both defect related and intrinsic breakdown. Of these the most commonly employed are the log normal and Weibull distributions. The latter is a form of extreme value statistics which was used successfully to describe the failure of cable insulation (149). The assumptions and properties of both log normal and extreme value statistics and their suitability to describe breakdown will be considered in 3A.3.2 and 3A.3.3. First, however, the mechanism of breakdown in dielectric strength and wearout measurements will be

discussed in the light of the observed failure distributions.

3A.3.1 Mechanism of breakdown.

Processing and measurement conditions often seem to have different or even opposite effects on wearout and fast voltage ramp measurements. This could perhaps be taken as an indication that a different mechanism leads to breakdown in the two cases. However, there is some statistical evidence to suggest that the same mechanism is operating in both cases. In support of this view, Van der Schoot and Wolters (37) cite the fact that for a given wafer the intersection of the defect related and intrinsic parts of the failure distribution occurs at the same value on the cumulative percentage scale in both measurements. Crook (14) found the same relationship of applied field to t_{bd} held in all wearout measurements, whether they lasted for microseconds or thousands of hours. This suggests that there was no change in the breakdown mechanism over this time range. The most convincing evidence that only one mechanism operates is possibly that of Berman (39). He took an empirical relationship between time and field sometimes used to extrapolate the results of wearout tests:

$$\ln t = b(F) - \gamma E_a \qquad (3.7)$$

E_a is the applied field, b(F) the failure distribution and t the time. Assuming a similar relationship could be applied to ramp tests if the time at a given field were weighted for the probability of failure at that field, he derived the following relationship between ramp and wearout tests:

$$\ln t(F) = \ln t_o + \gamma[E_b(F) - E_a] \qquad (3.8)$$

where ln t(F) is the wearout distribution, t_o, the effective time at field ($t_o = 1/\gamma R$ where R is the ramp rate), $E_b(F)$ is the breakdown field distribution and E_a the applied field in the wearout test. Having determined the constant γ experimentally, equation (3.8) could

be used to predict wearout distributions from ramp results and vice versa with considerable success.

3A.3.2 Extreme value statistics.

De Wit et al. (150) and Solomon et al. (151) derived a form of extreme value distribution founded on three basic assumptions:

1. Breakdowns are random independent events i.e. there is a Poisson distribution over area and time.
2. There is a constant a priori probability of breakdown at a constant field.
3. The mean time between breakdowns decreases exponentially with applied field.

They obtained a probability density function of the form:

$$\rho(t) = \beta\dot{E}\ \exp\{-y-\exp(-y)\} \qquad (3.9)$$

where $-y = \beta\dot{E}t - \ln(\beta\dot{E}t_{bo})$, t is the time, $\dot{E}$ the ramp-rate and β and t_{bo} are constants. This probability function has a strong negative skew which is also found experimentally for high field breakdown distributions. The agreement with experiment was only qualitative for SiO_2 however. Shatzkes, Av-Ron and Gdula (105) extended the model to include the effect of defects and by an appropriate choice of parameters were able to obtain a reasonable fit to the data of Osburn and Ormond (5) for a sample showing a distribution with a significant proportion of medium field breakdowns.

The Weibull distribution is the most commonly used form of extreme value statistics. Hill and Dissado attempted to correlate the Weibull time and field parameters with known properties of the materials (152). They used a model of dielectric relaxation response based on the existence of correlated regions in the material (153). Time dependent fluctuations occur due to dipolar exchange processes

between these correlated regions. This can cause electrical noise (154) of the form $1/f^{\alpha}$ where f is the frequency and α is related to the correlation index for synchronous dipolar exchanges, m, by $\alpha = 2(1-m)$. The fluctuations can be considered as random self-cancelling eddy currents. In a medium of finite resistivity these give rise to local voltage fluctuations. In the presence of an applied field fluctuations in the polarisation can result in very high instantaneous local fields which may cause breakdown.

The breakdown probability at an applied field E_a and time t was derived as:

$$1 - \exp[-N\beta^{1/m}[(E_a-E_1)/E_o)]^{1/m}] \qquad (3.10)$$

β is a geometrical factor, E_1 and E_o are electrical fields characteristic of the material and N is the number of correlated fluctuation elements in the sample.

The breakdown distribution was derived for three conditions: homogeneous breakdown occurring immediately a critical field is exceeded in a single fluctuation element, breakdown initiated by partial tree growth and breakdown caused by tree growth through the thickness of the sample. In all cases they obtained equations of the form:

$$F(E,t) = 1 - \exp\{-C(t/t_o)^a[(E_a-E_1)/E_o]^b\} \qquad (3.11)$$

where F(E,t) is the cumulative survival probability, a, b and C are constants comprised of varying parameters in each of the three cases and t_o is a scaling factor for the time. It can be seen that equation (5,13) is very similar to the two parameter Weibull function which has been found to give a good fit to data from static tests (155).

$$F = 1 - \exp(-C't^aE^b) \qquad (3.12)$$

C' now contains the scaling parameters of field and time and E_1 is assumed to be very small and is neglected. The survival probability defined by the above equation can be considered as the product of survival probabilities for each individual time increment upto time t:

$$F = 1 - \exp(-C \int_0^t a t^{(a-1)} E^b . \, dt) \qquad (3.13)$$

Integration allows the corresponding equation for dynamic tests to be obtained (152). Assuming a linear ramp-rate $\dot{E}$, the instantaneous field at time t is E_t and

$$F = 1 - \exp(-Ca/(a+b).t^{(a+b)}\dot{E}^b) \qquad (3.14)$$

Wolters has also used this distribution for both wearout and dielectric strength measurements (32). However he chose a different relationship of field to time-to-breakdown, namely, E proportional to ln t rather than t proportional to E^{-n} (where n = b/a) as used by Hill and Dissado. ln(-ln(1-F)) is then plotted versus E rather than ln E in the Weibull plot. It is difficult to distinguish which time-field relationship best fits the measured data. Possibly different relationships hold for different experimental conditions since Mason (157) observed different field functions for the time to initiation of tree growth in insulators depending on the sample thickness.

For defect related breakdown (<9 MV/cm) the area dependence was found by Wolters to be compatible with the stability postulate of extreme value statistics and hence with a random distribution of defects over the capacitor (32). (The stability postulate states that the smallest or largest values of a group of observations are distributed in the same way as the original distribution if the original distribution itself consists of smallest or largest (extreme) values.)

3A.3.3 Log normal statistics.

Several workers have used log normal statistics to obtain an empirical fit to their data (14,38). Metzler (158) provided a theoretical justification for its use based on the following assumptions:

1. The probability of failure per unit time is proportional to the current density.
2. The oxide current is due to F-N tunneling.
3. Electron trapping induces a time (and current) dependent field distortion distributed randomly over the capacitor.

He needed also to introduce a 'field enhancement factor' to obtain realistic results.

The log normal and Weibull distributions can appear rather similar and it is difficult to distinguish them by curve fitting procedures. However, Wolters has shown using the wearout data of Crook (14) that integration of the failure rate implies an extreme value rather than a log normal distribution. Berman has also indicated that it is actually not possible for the log normal assumption to be strictly obeyed for TDDB. Even if some populations were normal, an inhomogeneous population composed of devices drawn from two or more such populations would not itself be normal. Furthermore, for dielectric breakdown occurring at defect sites if the distribution of a particular population were normal then for random defects the distribution for a population differing only in device size would not be normal since the fractional loss would transform according to the Poisson relation.

3B: RESULTS AND DISCUSSION

3B.1 Charge trapping preceding breakdown.

Dielectric strength and wearout measurements have been the most used methods to date for studying breakdown. They provide useful information concerning the effects of technology on oxide integrity (17,19,111), allow the estimation of defect densities and also of the expected lifetime of the device in the field. However, they provide only indirect information on the actual mechanism by which breakdown occurs. Since most models for breakdown of SiO_2 are based on the build-up of trapped charge or the generation of traps, more direct evidence on the validity of these models can perhaps be obtained by examination of the charging phenomena occurring under the high field stress conditions which exist prior to breakdown. Harari (30) and Solomon (71) have both used C-V and ramp I-V measurements to try and do this and they inferred the presence of bulk and interface traps respectively. Quantitative analysis of I-V data is difficult, however and the high field required to achieve F-N tunneling may disturb the existing charge distribution. In this work, ramp I-V curves were used only to obtain a qualitative picture of charge trapping. A more accurate technique, photo I-V, was used to measure the charge density and to locate the centroid of oxide charge distributions. Constant current high field stress was used to simulate the conditions existing prior to breakdown and avalanche injection to fill traps in the oxide.

3B.1.1 Ramp I-V measurements.

A typical ramp I-V curve has already been shown in Fig.1.5. This is reproduced in Fig.3.7 for convenience and in Fig.3.8 is a Fowler-Nordheim plot of the same measurement showing that the F-N tunneling equation indeed holds over several orders of magnitude of the current density. At high fields the I-V curve begins to level off (Fig.3.7), indicative of reduced electron injection, probably caused by electron trapping in the oxide (71). Finally breakdown occurs.

The current at which the ledge is found depends on the ramp rate (111). If the end of the ledge is reached before breakdown, the width and current density at which it occurs may be used to obtain a rough estimate of the charge density and the capture cross section of the electron traps (71). Frequently breakdown occurs before the end of the ledge is reached however.

If the I-V measurement is stopped before the point of breakdown and then immediately repeated the presence of both positive and negative charge generated during the first ramp can be seen (Fig.3.9, curve 2). The early onset of conduction is indicative of a reduced barrier height due to positive oxide charge. The later shift of the curve to higher voltages indicates that there is also a region of negative charge density i.e. trapped electrons in the oxide. If a positive bias is applied between taking the two I-V curves no positive charge is seen but merely a lateral shift of the I-V curve due to negative charge (Fig.3.9, curve 3). This behaviour of the positive charge is consistent with its being due to slow interface states. Slow-states were first observed during avalanche injection of electrons. They are created close to the $Si-SiO_2$ interface in a positively charged condition and can be neutralised by moderate positive biases or elevated temperatures and charged again under negative bias.

3B.1.2 ΔV_{fb} during high field stress.

As a result of charge trapping in the oxide a shift occurs in the high frequency C-V curve following high field stress. Although during such a stress electrons are being injected by F-N tunneling, a shift to negative voltages was frequently observed indicative of net positive oxide charge. This is illustrated in Fig.3.10, curve (b), for a 52 nm oxide on a p-type substrate after injection of 9.76×10^{-3} C/cm^2 of charge (2×10^{-7} A/cm^2 for 48800 s). Avalanche injection of electrons, on the other hand, results in electron trapping and a positive shift of the C-V curve. Curve (c) in Fig.3.10

shows such a shift after avalanche injection at 4×10^{-6} A/cm^2 for 15000 s i.e. 6×10^{-2} C/cm^2 of injected charge. The lower field at which current can be injected by the latter technique (~4 MV/cm) favours electron trap filling whereas the high fields needed for F-N tunneling (9-10 MV/cm) cause a reduction in the capture cross section of the electron traps and favour field ionisation. Avalanche injection after a high field stress fills these electron traps again and net negative oxide charge is again seen (Fig.3.10, curve (d)). It is clear from Fig.3.10 that by monitoring a specific point on the C-V curve, such as the flatband voltage, qualitative information can be obtained about the charge build-up under constant current high field stress or indeed during avalanche injection. Typical curves showing the flatband voltage shift, ΔV_{fb}, as a function of time for n- and p-type samples at both positive and negative voltages are shown in Fig.3.11. The injection current density was 2×10^{-3} A/cm^2 in each case. Again, it is notable that initially quite large amounts of positive charge were generated in all cases (negative ΔV_{fb}) despite the fact that it is electrons that are being injected.

At positive voltages the positive charge is eventually neutralised and/or it is compensated for by electron trapping (ΔV_{fb} becomes positive). This is most clearly seen for the n-type sample stressed in accumulation (Fig.3.11 (a)). For the p-type sample, stressed in inversion, this neutralization/compensation process was only seen when the sample was illuminated (Fig.3.11 (c)). The slow generation rate of minority carriers in the lowly doped substrate used here is most likely responsible for the failure to observe any electron trapping (positive ΔV_{fb}) unless light is applied. If, in the dark, all or most of the generated minority carriers are required to maintain the set current, not enough carriers will be available to maintain the inversion layer. The surface potential will then differ from that when the light is switched on and the surface is kept in equilibrium by the enhanced minority carrier generation rate. The surface potential can in turn affect the neutralisation of positive charge and also the rate of electron trapping.

At negative voltages only net positive charge is observed (Fig.3.11 (b) and (d)). To a first approximation the polarity of the field rather than the type of substrate seems to be the most important factor in determining ΔV_{fb}. Hence to avoid problems in generating the inversion layer, measurements were usually carried out in accumulation on p- and n-type substrates to examine the effects of negative and positive voltage stresses respectively.

The relative rate of positive and negative charge build-up for an n-type sample stressed in accumulation depends on the current density used for injection. Figure 3.12 shows ΔV_{fb} versus injected charge curves for a 39 nm oxide stressed at positive gate voltages and current levels of 2×10^{-3}, 2×10^{-4} and 2×10^{-5} A/cm^2. In all cases, any positive charge that was generated appeared at the beginning of the stress. This was subsequently neutralised and/or compensated for by electron trapping as injection continued. No net positive charge was seen at the lowest current density (field). The increase in the initial negative ΔV_{fb} with current density means that either more positive charge is generated and/or the rates of positive charge neutralisation and electron trapping decrease with increasing current density. The latter is most likely since it is known that the capture cross section for electron trapping is field (and hence current) dependent. If the positive charge is indeed due to slow states, these are also known to be more effectively neutralised at moderate than at high fields i.e. at lower current levels.

3B.1.3 Photo I-V measurements.

To see whether the positive charge created is really located at the interface and is due therefore to slow states, photo I-V measurements were performed before and after a high field constant current stress at negative gate voltage. This polarity was used to avoid any neutralisation of the positive charge. After a stress on a p-type sample of 2.65×10^{-4} A/cm^2 for 2000 s (or 0.53 C/cm^2 of

injected charge), there was a flatband voltage shift of more than 11 V indicative of substantial positive charge generation. This shift is shown in Fig.3.13. However, there is no shift at all in the photo I-V curve taken after stress relative to the initial curve (Fig.3.14, curve 2). This means that the positive charge is located in the region to which the photo I-V technique is not sensitive i.e. within about 10 Å of the $Si-SiO_2$ interface. Therefore the positive charge is indeed interface rather than bulk charge.

It is interesting to note that there is also no evidence of bulk electron trapping following the high field stress. The slight deviation of photo I-V curve 2 from curve 1 in Fig.3.14 at the higher positive values of the applied field (injection from the silicon) suggests that there is a small amount of electron trapping within about 30 Å of the $Si-SiO_2$ interface. When avalanche injection of electrons was carried out after the high field stress (2.6×10^{-4} A/cm^2 for 200 s) there was a positive shift of the flatband voltage indicative of electron trap filling (Fig.3.15). From Fig.3.14 it is evident that this is bulk charge since there is a parallel shift of the photo I-V curve along the voltage axis, relative to the initial curve, after avalanche injection.

Recent work, also using the photo I-V technique and capacitors with both aluminium and polysilicon gate electrodes has shown that electron traps are not only filled but are created at the non-injecting interface during high field stress (146). (More traps could be filled by avalanche electron injection after a high field stress than on an unstressed sample. Hence traps must have been created during the stress.) Only those energetically deep traps close to the $Si-SiO_2$ interface are actually filled under high field stress conditions, however, since high fields favour detrapping of charge. The centroid of the traps created during high field stress was calculated to be at about 50 Å from the $Si-SiO_2$ interface. Thus, these defects are generated mainly near to the anode.

Photo I-V measurements were also performed before and after constant current high field stress at positive gate voltages i.e. F-N tunneling from the silicon substrate. (From the appreciable negative ΔV_{fb} after such a positive constant current stress it was again concluded that slow interface states were present.) Figure 3B.16 shows the photo I-V curves obtained on a p-type sample before and after a stress of 2.65×10^{-5} A/cm^2 for 1000 s or 2.65×10^{-2} C/cm^2 of injected charge. After stress at this polarity there is some electron trapping near to the Si-SiO_2 interface. This is indicated by the shift to higher voltages relative to the initial curve of the photo I-V curve for injection from the substrate (positive gate voltage). For these aluminium gate devices no electron trap generation could be seen at the aluminium-SiO_2 interface i.e. at the non-injecting interface. Instead positive charge appears to be generated here. This can be seen from the increase in the photocurrent for a given negative applied voltage (i.e. photoinjection from the gate) relative to the initial curve. Experiments using transparent polysilicon gates have demonstrated that this positive charge is related to the presence of an aluminium gate electrode (146,159). With polysilicon gates, electron trap generation is seen instead under these injection conditions. Avalanche injection of electrons after the positive constant current stress could not completely neutralise the aluminium-related charge. This can be seen from the photo I-V curve taken after avalanche injection for 200 s at 2.65×10^{-4} A/cm^2 also presented in Fig.3.16. The results presented are for p-type substrates but comparable data were obtained with n-type wafers.

3B.1.4 Model for charge build-up.

From the above observations the following model for charge trapping and defect creation during high field stressing was evolved. This is also summarised in Fig.3.17.

1. Positively charged 'slow-states' are created at the Si-SiO_2 interface (<10 Å) at both stress polarities.

2. Electron traps are created at the non-injecting interface (anode).
3. Only a few of the energetically deep traps close to the interface (<30 Å) are filled under negative high field stress conditions.
4. Electron traps near to the $Si-SiO_2$ interface are filled during positive high field stress. (A different mechanism for trap ionisation is probably operating at this polarity. Under positive stress, trapped electrons must tunnel to the SiO_2 conduction band in order to leave the oxide. In this case the trap depth is the determining factor. At negative polarity the electrons must tunnel to the silicon. In this case only the distance from the interface is important.)
5. Positive charge is generated at the gate under positive bias conditions when aluminium gate electrodes are used.
6. No bulk charge trapping was seen for high field stress under any injection conditions.

3B.1.5 Discussion of breakdown models.

This picture of charge trapping in the oxide under high field stress has several implications for the breakdown mechanism. Firstly, electron trap creation at the anode suggests that this region is where the most damage is occurring. It is not surprising that this is the case since the electrons attain their maximum energy at the non-injecting interface. (Considerable electron heating at quite low fields has been demonstrated and the electron energy is well correlated with the anode field (18). This trap generation is compatible with the electrochemical models for breakdown based on the growth of damage 'trees' between the electrodes (31,32,130,131). Harari's model also stresses the role of created electron traps in causing breakdown (8). However, from his data he inferred the generation of such traps at the cathode. No evidence for this was found from the photo I-V measurements. Harari's oxides were extremely

thin though, (44-160 A) and the distinction between 'near to the cathode' and 'near to the anode' is hard to make. Nevertheless, the ledge that is observed in the I-V curve at high fields means that something is occurring to reduce electron injection. This could either be electron trapping near enough to the cathode to result in a reduction of the injecting field and hence of the current, or defect generation causing reduced injection efficiency.

The resonant tunneling model also emphasises the role of defects near the cathode in enhancing the current and causing breakdown (132). Under constant current conditions there is an increase in the applied voltage (and hence the cathode field) with time which could allow the resonant condition to be achieved. Under constant voltage conditions, however, it is difficult to see how this mechanism could lead to breakdown without the generation of new traps near to the cathode for which no evidence was found.

The impact ionisation model requires a build-up of positive charge at the cathode to enhance electron injection (6,134). Although positive charge was observed at the $Si-SiO_2$ interface this is in fact probably too close to the interface to have any effect on the injection current, since it is located before the barrier maximum. Along with its inapplicability to very thin oxides this makes the impact ionisation model seem an unlikely explanation of the breakdown event. It is not impossible that the positive charge does play a role in causing breakdown though. Fischetti has suggested that at anode fields above ~5 MV/cm hot holes are injected into the SiO_2 from the anode as a result of surface plasmon decay and that these holes are responsible for the observed positive charge creation (144). Both positive charge build-up during high field stress (164) and the breakdown event (165,166) show a correlation with the anode field. It is possible therefore that injected hot holes cause or contribute to damage creation at the anode and thereby contribute to breakdown.

3B.2 Charge to breakdown under constant current stress.

The effect of current injection conditions and sample history on Q_{bd}, the total amount of charge which can be injected prior to breakdown, should also provide some useful information regarding the breakdown mechanism (72,73,111). In particular, the relative influence of current and field on the point of breakdown is still under dispute. Some authors contend that the parameter which essentially determines this is the total number of injected carriers. They propose that at least to a first approximation Q_{bd} is a constant (33,36,160), independent of the injection rate (current density) and hence, for F-N tunneling, of the field (energy of the injected electrons). Evidence was recently provided to the contrary, however. Experiments on transistors were able to separate the effects of field and current on Q_{bd} and this parameter was found to be strongly field dependent, although little affected by the rate of injection (161). In the remainder of this section the effect of current density, duty cycle, electrode material and temperature on Q_{bd}, measured by constant current stressing of MOS capacitors, will be examined in more detail.

3B.2.1 Variation of Q_{bd} with current density.

From Fig.3.12, showing ΔV_{fb} as a function of injected charge for stresses at different current densities, it already appears that there is a decrease in Q_{bd} with increasing current density i.e. with increasing field. (Care was taken to ensure that the flatband voltage was measured at the same intervals of injected charge at each current level so that there should be no contribution from any relaxation effects to any change in the point of breakdown.) Further indication that the injection field is of importance comes from a comparison of Q_{bd} measured at the same current density under F-N tunneling and avalanche electron injection conditions. Values of Q_{bd} at two set current levels and using both means of injection are shown in Table 3.3 for wafer P512. There is a factor of about 2 to

2.5 difference in the field applied in the two techniques (around 4 MV/cm for avalanche and 9 - 10 MV/cm for F-N injection). More charge could be injected at both current levels under the lower field avalanche conditions without causing breakdown.

The results in Table 3.3 are only for one randomly chosen capacitor at each set current level and each injection method. More systematic measurements of Q_{bd} as a function of current density (and hence of field) were carried out using the HPK220 constant current source rather than the avalanche injection system. Injection was only by F-N tunneling therefore but a wider range of current densities could be examined. Several measurements were also made at each set current to obtain more statistically significant results. A minimum of five capacitors of area $4.91 \times 10^{-4} cm^2$ were measured at each current density. Both the maximum value of Q_{bd}, $Q_{bd}(max)$, and the average value for the high field breakdowns, $\overline{Q_{bd}}$ were measured. A plot of $Q_{bd}(max)$ and $\overline{Q_{bd}}$ versus injection current density is shown in Fig.3.18 and both parameters show the same trend. There is indeed a range of two to three orders of magnitude (5×10^{-3} to 2 A/cm^2) over which Q_{bd} appears to vary relatively little. At very high current levels (>0.5 A/cm^2), Q_{bd} could perhaps be considered to fall off slightly more rapidly. Wolters has observed a decrease at high injection currents and has suggested that it is caused by inhomogeneous injection (37). However, it is also clear from Fig.3.18 that Q_{bd} increases quite significantly at low current levels, contradicting the view that it remains essentially constant regardless of injection conditions.

From their measurements on transistors, where they were able to vary the oxide field and injection current independently, Modelli and Ricco (161) found Q_{bd} to be exponentially dependent on the oxide field such that:

$$Q_{bd} = Q_o \exp(-\alpha/E_{ox}) \qquad (3.1)$$

where Q_o and α are constants equal to 1.6×10^{-2} C/cm^2 and 8.1×10^7 V/cm[1] respectively at 300 K. Provided that the field across the oxide was kept constant, they found Q_{bd} to be almost independent of the current density used in their measurements. On capacitors it is not possible to vary the current density independentiy of the applied field, since these parameters are related by the F-N tunneling equation. Hence the results of reference (161) would predict a variation of Q_{bd} with the current density due to the accompanying change in applied field.

In a constant current stress measurement on an MOS capacitor the applied voltage varies to compensate for charge trapping in the oxide. The cathode field is kept constant by this process. As charge trapping has been shown to be very low at high fields this approximates to keeping the net oxide field constant. In an attempt to apply equation (3.1) to constant current measurements the average value of the initial applied field at each current density was taken as the value of the oxide field, E_{ox}. Extraction of E_{ox} from a measurement of the applied field as a function of injected charge is shown in Fig. 3.19. Negligible charge trapping has taken place at the point of measuring E_{ox} and it is assumed that all subsequent changes in the applied field are to adjust the cathode field back to this initial value. A plot of $\ln Q_{bd}(max)$ versus $1/E_{ox}$ can then be made and this is shown in Fig.3.20. There is a certain amount of scatter on the data due partly to error in the measurement of E_{ox} at low set current levels and partly to statistical variation in Q_{bd}. A reasonable straight line fit can be made however and values of the intercept and gradient are in excellent agreement with those from ref. (161): $Q_o = 1.74 \times 10^{-2}$ C/cm^2 and $\alpha = 8.07 \times 10^7$ V/cm[1]. The field dependence of Q_{bd} is therefore confirmed. However, it can <u>appear</u> that Q_{bd} is constant over quite a wide range of current densities for constant current stress measurements on capacitors because of the exponential dependence of the current on the field in F-N tunneling. This means that only a small change in field is required to produce several orders of magnitude change in the current.

Hence Q_{bd} varies strongly with current density only at low set currents.

It can be concluded from these results that both current and field are required to achieve the breakdown condition. The importance of the field and hence of the electron energy can be clearly seen from Figure 3.20. However, the constant value of Q_{bd} with varying current density provided the field is kept constant (161) suggests that a certain amount of damage must indeed be done by the incoming electrons before breakdown occurs. It has also been shown that very high fields (30 MV/cm) may be applied to the oxide for extremely short times (1 μs) without causing breakdown (162). Under these pulsed voltage conditions the amount of current injected is relatively low and the value of Q_{bd} required for breakdown is not reached despite the fact that this is also low at such high fields.

3B.2.2 Dependence of Q_{bd} on duty cycle.

Since larger values of Q_{bd} resulted from stressing at lower currents (fields) where there was least positive charge build-up (Fig.3.12), an experiment was conducted to see whether or not the charge state of the slow-states had any influence on breakdown. On stressing wafer L9 continuously at +2 $mAcm^2$ (positive gate voltage) breakdown consistently occurred after injection of about 0.5 C/cm^2. If a moderate positive voltage (~5 MV/cm) was periodically interspersed with the constant current stress to neutralize the slow-states a two to three times increase in Q_{bd} occurred. This is illustrated in Fig.3.21. The increase in Q_{bd} could be caused by the charge state of the slow-states or simply be due to the periodic relaxation of the voltage. Whichever is the case, Q_{bd} is again shown not to be independent of the injection conditions.

Further measurements on the effect of interrupting the constant current stress or interspersing a different stress were conducted on another batch of wafers using the Keithley 220 current source rather

than the avalanche system. This allowed more accurate control of the stress times and a wider range of measurement times and currents. For wafer ST2, stressed at a constant current density of +2.04 mA/cm^2, the maximum Q_{bd} was 1.02 C/cm^2. If, at intervals of 25 secs, this stress was interrupted for 5 secs and a lower positive or a negative voltage applied instead Q_{bd}(max) increased. When 1 V was interspersed during the positive constant current stress, the maximum value of Q_{bd} increased to 1.8 C/cm^2. -10 V as the interspersed stress caused a further increase in Q_{bd}(max) to 2.3 C/cm^2. Most surprisingly, interspersing a negative stress sufficient to cause current injection at -2 mA/cm^2 (approximately -30 V) increased Q_{bd}(max) (at positive polarity alone) to 4.9 C/cm^2. Fig.3.22 shows these results plotted as the variation in Q_{bd}(max) and also in $\overline{Q_{bd}}$ with the magnitude of the interspersed voltage. Where a constant current was used rather than a constant voltage as the interrupting stress an approximate value of the voltage has been chosen.

Increasing the frequency of the interruption also caused an increase in Q_{bd}. This is shown in Figs.3.23 and 3.24 where Q_{bd}(max) and $\overline{Q_{bd}}$ at +2 mA/cm^2 are plotted as a function of the interrupt interval for interspersed stresses of -10 V and -2 mA/cm^2 (--30 V) respectively. The change in Q_{bd} with interrupt interval was greater for the more negative interspersed stress. Varying the length of the interspersed stress between 1 and 20 secs had very little effect on Q_{bd}, however.

These results appear to contradict those of reference (72) where it was found that a constant amount of charge could be injected at one polarity regardless of interruption or reversal of the stress. Most probably the single stress interruption used in ref. (72) on devices already stressed to 80% of the expected Q_{bd} was insufficient to affect the final value of Q_{bd}. This is to be expected from interpolation of Fig.3.24. Interrupting the stress at 80% Q_{bd}(max) or around 400 secs would have negligible effect on the measured value of Q_{bd}. Further experiments are needed to elucidate precisely what

is responsible for the increase in Q_{bd} on interruption of the stress or reversal of its polarity. It is apparently a relaxation effect however, rather than anything to do with the charge state of the slow-states, since both negative and low positive voltages enhance Q_{bd} and the former charges the slow-states while the latter discharges them.

3B.2.3 Electrode dependence of Q_{bd}.

Fig.3.25 shows ΔV_{fb} versus injected charge curves for polysilicon and aluminium gate devices at both polarities and using a current density of 2×10^{-3} A/cm^{-2}. Two differences between the aluminium gate and polysilicon gate devices are striking. Firstly, there is relatively much less positive charge build-up in the polysilicon capacitors. Secondly, Q_{bd} is over two orders of magnitude greater for the polysilicon gate devices. For injection from the gate the higher value of Q_{bd} for polysilicon could perhaps be partly explained by the lower injection field required (lower barrier height). For injection from the substrate the field needed to achieve a given current will be the same in both cases. However, in this case the region of most damage will be near the gate and the Al-SiO_2 interface is known to have more defects than the Si-SiO_2 interface. Perhaps at this polarity the anode region is more easily damaged when there is an aluminium gate, thereby reducing Q_{bd}.

As far as the difference in positive charge generation is concerned, it has already been mentioned that during high field stress at positive polarity there is considerable generation of metal-related positive charge with aluminium gate electrodes that is not found for polysilicon gates. At negative polarity, the difference is possibly related to the different injection field required to achieve a given current for the two electrodes. As shown in Fig.3.12, a lower field also results in relatively less positive charge generation at the Si-SiO_2 interface. This may be at least part of the reason for the lower positive charge seen in the polysilicon gate devices. Whatever

the reason for the reduced positive charge in polysilicon gate devices, since there is a correlation between amount of positive charge and Q_{bd} this provides circumstantial evidence in favour of the positive charge playing a role in determining the point of breakdown.

3B.2.4 Temperature dependence of Q_{bd}.

It was found that the average value of Q_{bd} was quite a strong function of the injection temperature. Table 3.4 shows the average value of Q_{bd} from a number of measurements at each of several temperatures from 20 to 80°C. The current density was 2×10^{-3} A/cm^{-2} and the wafer P512. There was more than an order of magnitude change in $\overline{Q_{bd}}$ over this range. Unfortunately it was not possible to conduct a statistical number of measurements at each temperature in order to obtain a more accurate picture of the effect of temperature on the failure distribution. The overall trend is, however, in agreement with that reported by Harari (8).

This very strong dependence of breakdown on temperature is rather difficult to explain. There is certainly no change in the hot electron distribution in the SiO_2 at these temperatures. Lattice vibrations will be enhanced though, increasing the probability of bond breaking by hot carriers, especially at defects or already weak bonds. Resonant tunneling is also expected to be enhanced by elevated temperatures (cf. 3A.2.3.4).

3B.2.5 Q_{bd} and breakdown models.

The dependence of Q_{bd} on the oxide field suggests that the energy as well as the number of the incoming electrons is important in causing breakdown. Both the anode hole injection (144) and the gas discharge model (130,131) are consistent with this observation since both defect generation and surface plasmon excitation at the anode depend on the electron energy at this point. The increase of Q_{bd} on

periodic interruption of the stress can be explained in the gas discharge model by resolidification of the gaseous channels on reduction of the voltage. Further charge injection is then required to recreate these channels. The fact that quantities of charge injected at positive and negative polarities do not simply add to give Q_{bd} is also consistent with either model based on damage occurring at the anode. On reversal of the polarity, the process of damage creation must presumably begin again at the new anode, thus postponing breakdown. Resonant tunneling (132) and the electron trap creation model of Harari (8) could also be expected to show a dependence on the applied field. However, the main objection to these models is that they require defect creation at the cathode, for which no evidence was found.

3B.3 Relationship of TDDB to dielectric strength.

As has been discussed previously, there is some evidence in the literature to suggest that breakdown occurs by the same mechanism in wearout and fast voltage ramp tests (14,36,39,160). Extreme value statistics and in particular the Weibull distribution, have also been shown to provide the best statistical description of the breakdown process from both experimental and theoretical considerations. In reference (155), the failure distributions under wearout and fast voltage ramp conditions are derived assuming a Weibull distribution for wearout of the form:

$$(1-F) = \exp(-Ct^{a}E^{b}) \qquad (3.2)$$

where F is the cumulative probability of failure, E the applied field, t the time and a, b and C are constants. The corresponding equation for a dielectric strength measurement provided that there is no change in the breakdown mechanism is then:

$$(1-F) = \exp(-[Ca/(a+b)]t^{(a+b)}\dot{E}^{b}) \qquad (3.3)$$

Here, $\dot{E}$ is the ramp-rate. If the above equations are applied to wearout and dielectric strength measurements on the same wafer and the parameters a and b are found to be the same in both cases it is extremely likely that the same breakdown mechanism is operating under both sets of test conditions. This analysis is described below for breakdown measurements on wafer T9.

3B.3.1 Applicability of Weibull distribution.

It was first necessary to ascertain that the above equations provided a good fit to the measured failure distributions. A typical breakdown histogram from a dielectric strength measurements at a ramp rate of 10 V/s is shown in Fig.3.26. Figure 3.27 contains the corresponding Weibull plot of ln(-ln(1-F)) versus ln V. Measurements were also carried out at ramp rates of 2.5, 25, 50 and 100 V/s. The histograms and Weibull plots for these measurements were essentially similar to those shown in Figs.3.26 and 3.27 apart from a shift of the breakdown distribution to lower field with decreasing ramp rate. Two distinct parts can be distinguished in the Weibull plots. The first linear portion with the smaller slope is due to defect related breakdown and the second steeper part represents 'intrinsic' breakdown. Only the latter was used for this analysis. Provided that these two regimes are considered separately, the fast voltage ramp measurement appears to be well described by the Weibull equation.

A Weibull plot for a wearout measurement conducted at an applied field of 10.5 MV/cm is shown in Fig.3.28. Similar measurements were also carried out at 8.5, 9.0, 9.25, 9.5, 9.75 and 10.0 MV/cm. All the Weibull plots were similar to that of Fig.3.28. A good straight line is found for ln(-ln(1-F)) versus ln t apart from some minor deviation in the early part of the distribution. This was again ignored as being due to defect-related breakdown.

3B.3.2. Analysis of wearout data.

For wearout measurements, a is obtained from the gradient of the Weibull plot. This was done for all the measurements at different values of the applied field and the resulting values are listed in Table 3.5. The average value of a is 0.24. Having obtained a, the value of b can be extracted from a plot of ln (1/t) versus ln E_a where E_a is the applied field and the time to breakdown t is taken at a specific value of F. The gradient of this plot is b/a. The value of F known as the estimator (F_e) was chosen as the characteristic point at which to measure t. The estimator is given by the condition that $Ct^aE^b = 1$, so that $F_e = 0.632$. Values of ln (1/t) and ln E_a are also given in Table 3.5 and plotted in Fig.3.29. From the gradient of this plot, b/a = 29.7, which for a = 0.24 gives b = 7.1.

3B.3.3 Analysis of dielectric strength data.

The gradient of the Weibull plot is in this case equal to (a+b). Values of (a+b) obtained from Fig.3.27 and analogous plots at the other ramp rates used are shown in Table 3.6. The average value is 7.46, if the anomalous data point at 50 V/s ramp-rate is ignored and 6.9, if all data points are taken into consideration. To determine a and b separately, ln (1/t) at F_e is plotted as a function of the logarithm of the ramp rate, ln $\dot{E}$. The data used for this plot are also listed in Table 3.6 and plotted in Fig.3.30. The gradient, which is equal to b/(a+b) was found to be 0.95. Hence taking (a+b) = 7.46, b = 7.1 and a = 0.36 or alternatively taking (a+b) = 6.9, b = 6.55 and a = 0.34. Either way, agreement between the two sets of data from wearout and dielectric strength measurements is good, suggesting that at least for applied fields in excess of 8.5 MV/cm breakdown proceeds by the same mechanism in wearout and fast voltage ramp measurements.

3B.3.4 Significance of the Weibull parameters.

From the value of the parameters a and b some deductions can be made about the nature of the breakdown process. When $a > 1$, the breakdown distribution in time is peaked. This implies an ageing process (155,163) in which the sample retains memory of any previous deterioration. In this case any stressing increases the probability of failure in a subsequent test. For $a = 1$, the number of breakdowns decays exponentially with time and the breakdown process is therefore perfectly random i.e. the sample retains no memory of prior testing. If $a < 1$, as we observe, the time distribution of breakdowns is not peaked and prior testing increases the reliability of the material i.e. there is a decreasing probability of breakdown with time. This was described by Wolters as a decreasing hazard rate (123) and he suggested that in the case of SiO_2 it is due to the observed decay of the injection current with time under constant applied voltage conditions.

The parameter b is material dependent and is inversely proportional to the fluctuation index, m, which is a measure of the degree of correlation of the motions that transport the local strain between regions (154). It determines the shape of the field distribution function (155). Since m and b are properties of the material itself they may be determined from other types of measurement such as the frequency dependence of the dielectric susceptibility or the mechanical compliance (155).

BREAKDOWN OF SiO_2: RESULTS

Current density (A/cm^2)	Q_{bd} (C/cm^2)	
	F-N Tunneling (9-10 MV/cm)	Avalanche e^- injection (~4 MV/cm)
7.0×10^{-5}	0.0028	> 0.133
3.5×10^{-5}	0.35	> 1.75

Table 3.3 Comparison of Q_{bd} under F-N tunneling and avalanche injection conditions at the same set current densities.

Temp. (°C)	Q_{bd} (C/cm^2)	Average Q_{bd} (C/cm^2)
20	20.8	
"	1.1	9.5
"	6.6	
50	1.2	
"	4.2	
"	>26.3	6.9
"	2.1	
"	0.8	
62	0.6	
"	6.0	2.1
"	0.3	
"	1.6	
81	0.2	
"	1.5	0.6
"	0.1	

Table 3.4 Q_{bd} as a function of measurement temperature for wafer P512 (29 nm oxide) at an injection current of 2 mA/cm^2.

E_a (MV/cm)	a	ln E_a	ln (1/t) (at F_e)
8.50	0.233	2.1401	7.2
9.00	0.195	2.1972	5.6
9.25	0.276	2.2246	5.9
9.50	0.215	2.2513	3.8
9.75	0.229	2.2773	3.7
10.00	0.310	2.3026	2.8
10.50	0.219	2.3514	1.3

Table 3.5 Data from wearout measurements made at various applied fields, E_a. Values of a are obtained from the slope of the Weibull plots, e.g. Fig.3.28 and ln t at F_e (63.2% failures) is extracted from the same data. ln t is plotted versus E_a in Fig. 3B.29 to obtain b/a.

$\dot{E}$ (V/s)	(a+b)	E_{ch} (at F_e) (MV/cm)	$\dot{E}/E_{ch}$	$\ln(\dot{E}/E_{ch})$
2.5	7.54	10.5	0.2381	-1.4351
10	7.63	11.5	0.8696	-0.1398
25	7.36	12.5	2.8000	1.0296
50	4.74	14.5	3.4483	1.2379
100	7.23	15.5	6.4516	1.8643

Table 3.6 Data from dielectric strength measurements. Values of (a+b) are obtained from the slope of the Weibull plots, e.g. Fig.3.27. The characteristic value of the field, E_{ch} was taken at F_e (63.2% failures). $\ln(\dot{E}/E_{ch})$ is plotted against $\dot{E}$ in Fig.3.30 to obtain b/(a+b).

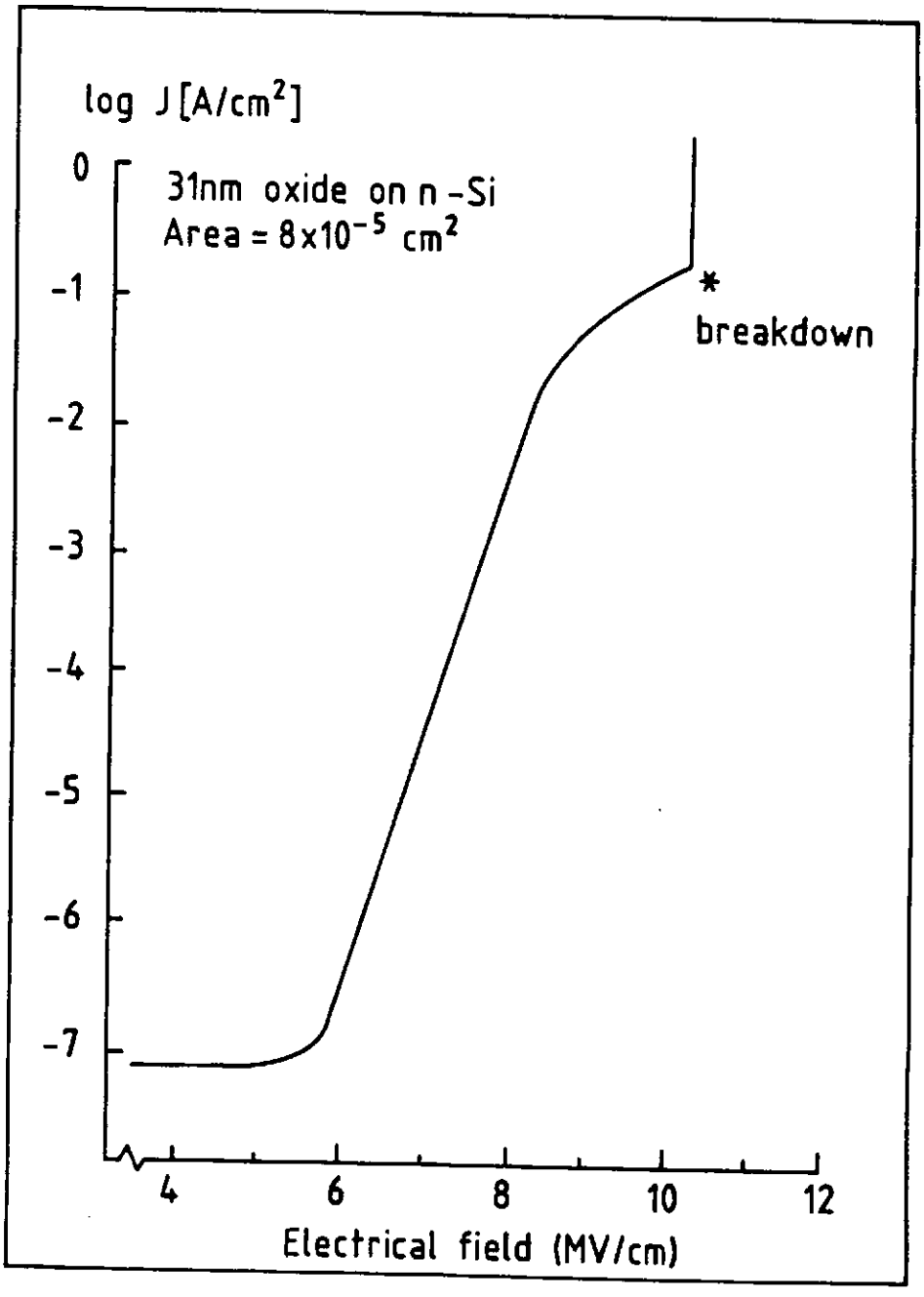


Figure 3.7 I-V curve measured on wafer JN1, cap J, at a sweep rate of 500mV/s.

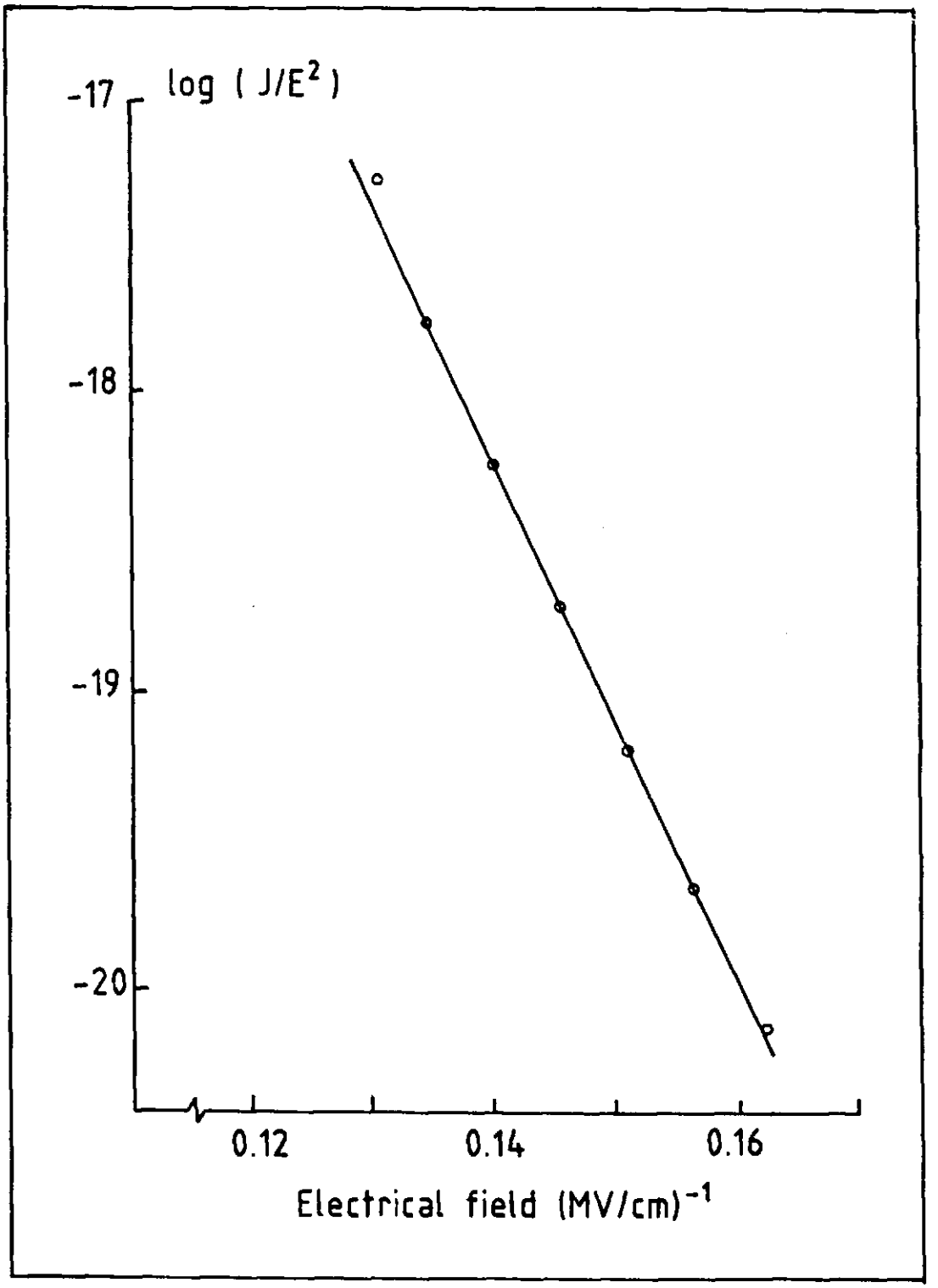


Figure 3.8 Fowler-Nordheim plot constructed from the I-V data contained in Fig.3.7. (wafer JN1 - cap.J)

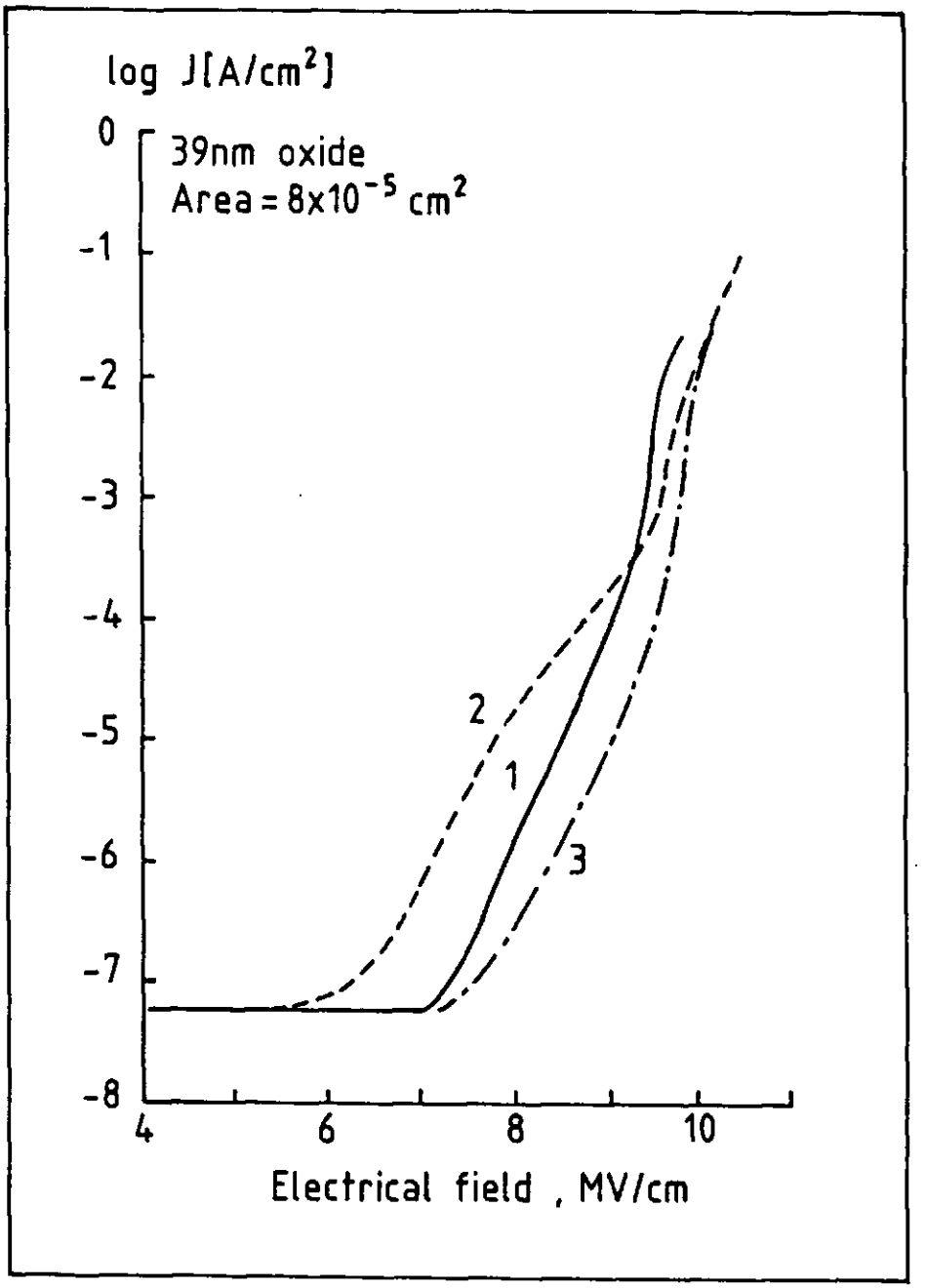


Figure 3.9 I-V curves measured on wafer L9, cap.J (1) first sweep to 9.5MV/cm. (2) second sweep to 10MV/cm. (3) After measuring curves (1) and (2) and subsequent application of a positive bias of about 5MV/cm.

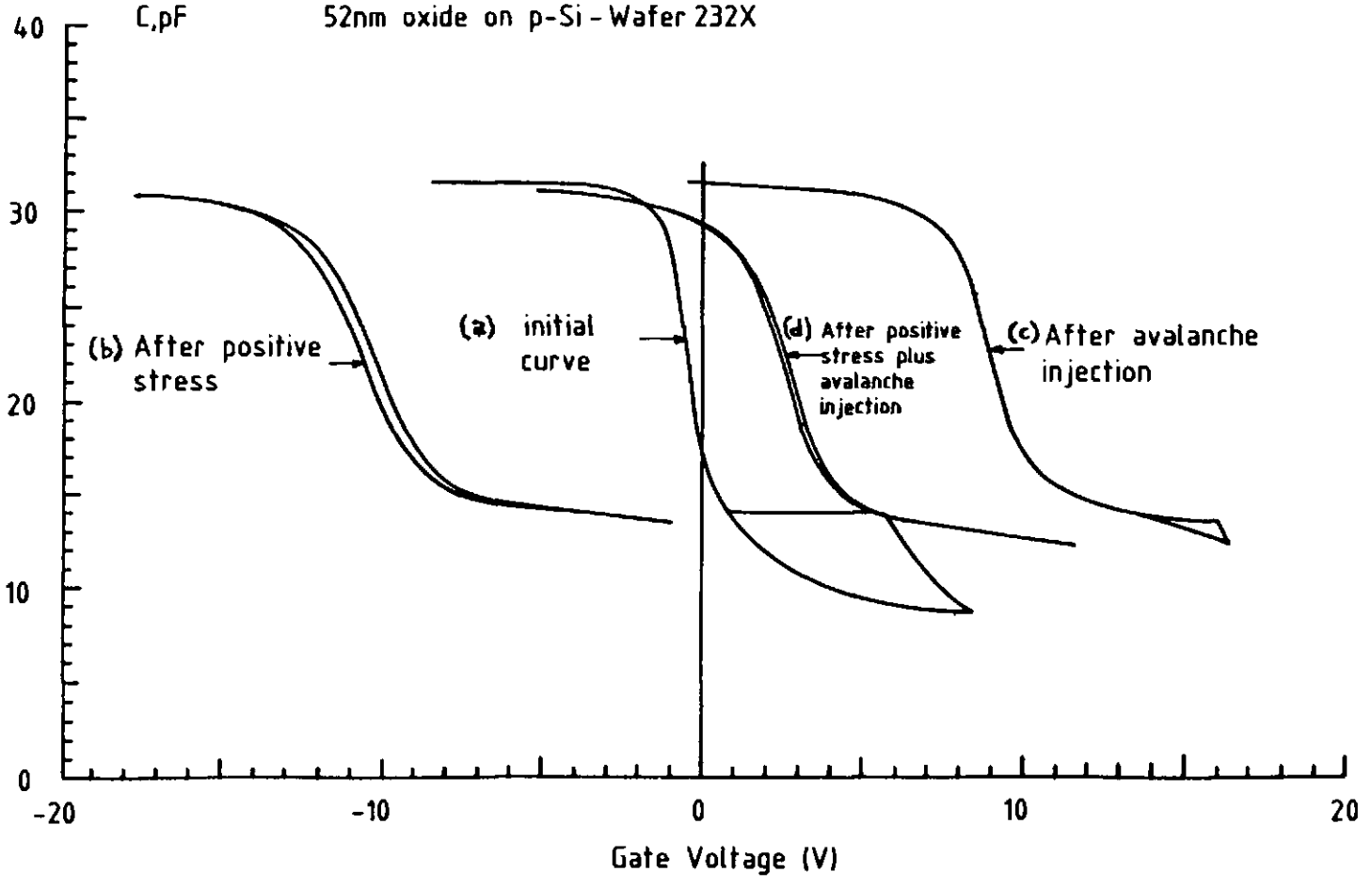


Figure 3.10 High frequency C-V curves (1MHz) taken after various stresses on cap.B, wafer 232X.

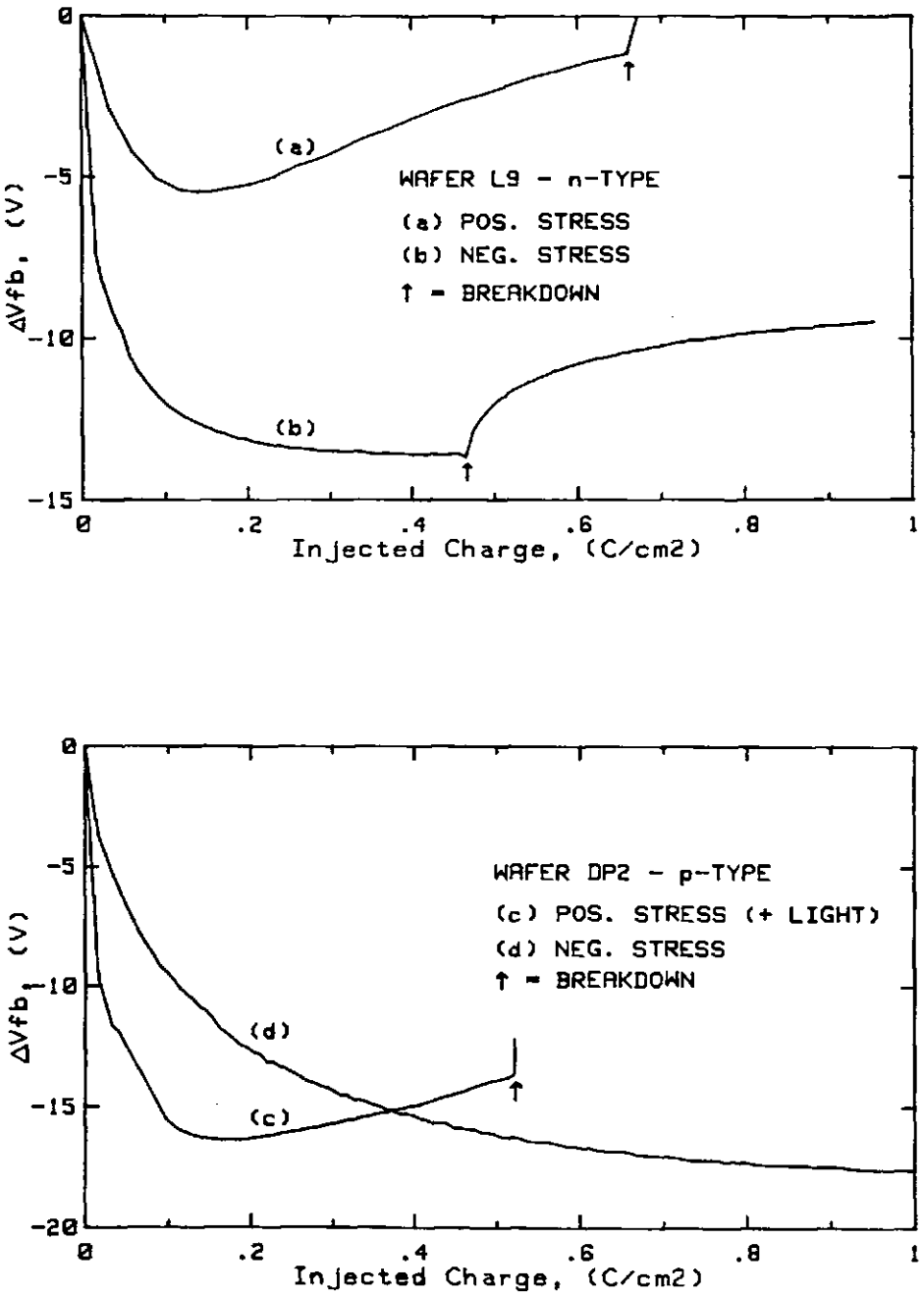


Figure 3.11 ΔV_{fb} as a function of injected charge for an n-type (L9) and a p-type (DP2) wafer stressed at $2mA/cm^2$.

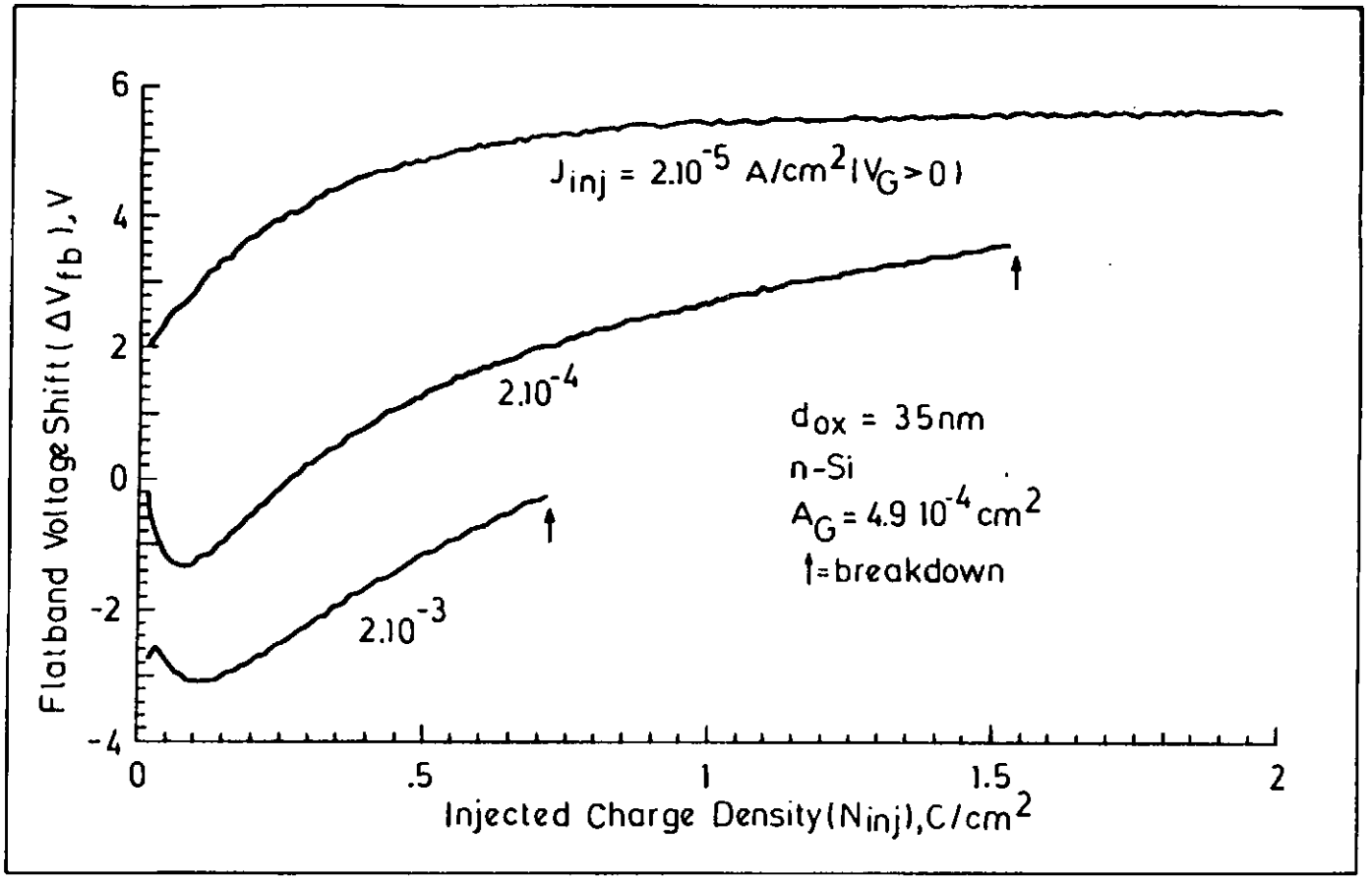


Figure 3.12 ΔV_{fb} as a function of injected charge for wafer L9, cap.B stressed at various current densities.

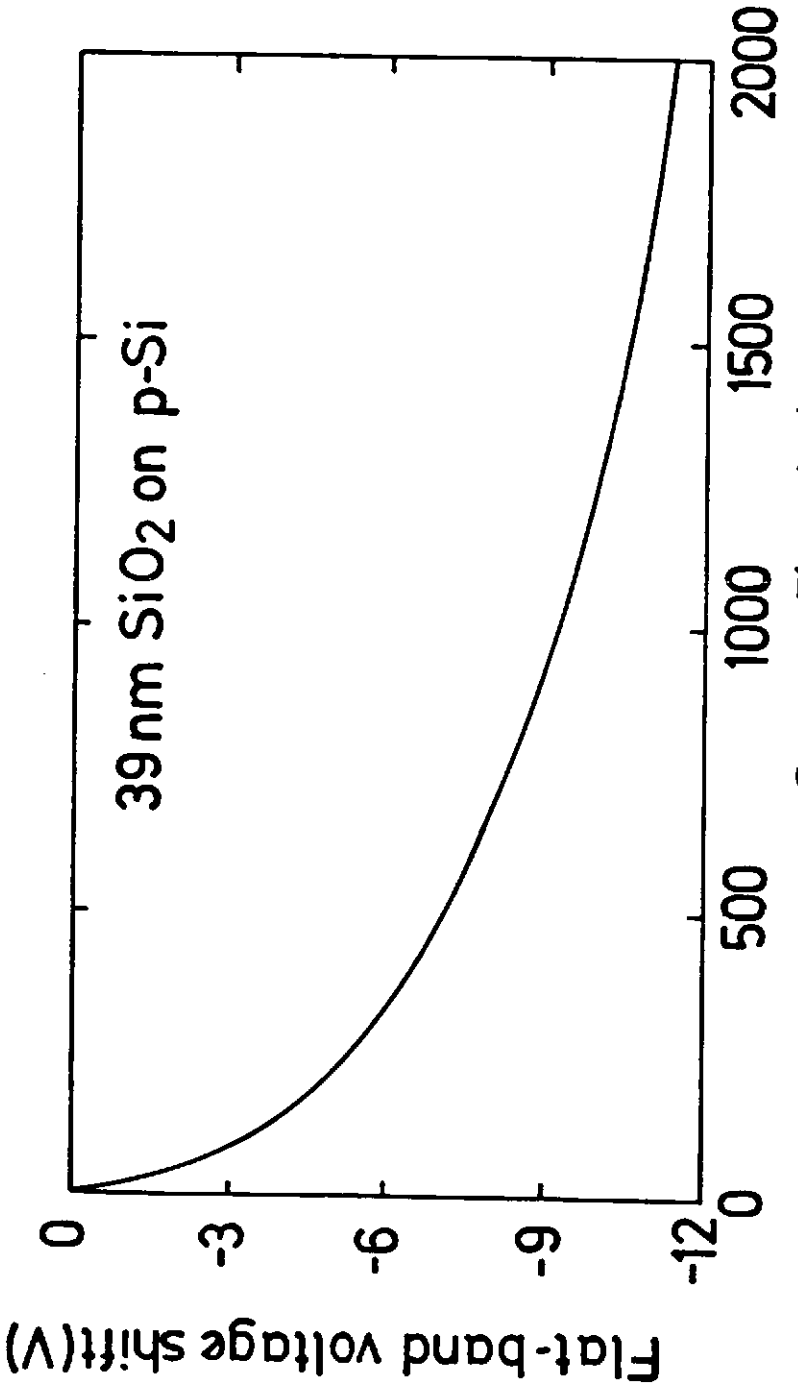


Figure 3.13 ΔV_{fb} as a function of time during high field stress at 2.65×10^{-4} A/cm^2 on cap.H, wafer MP4.

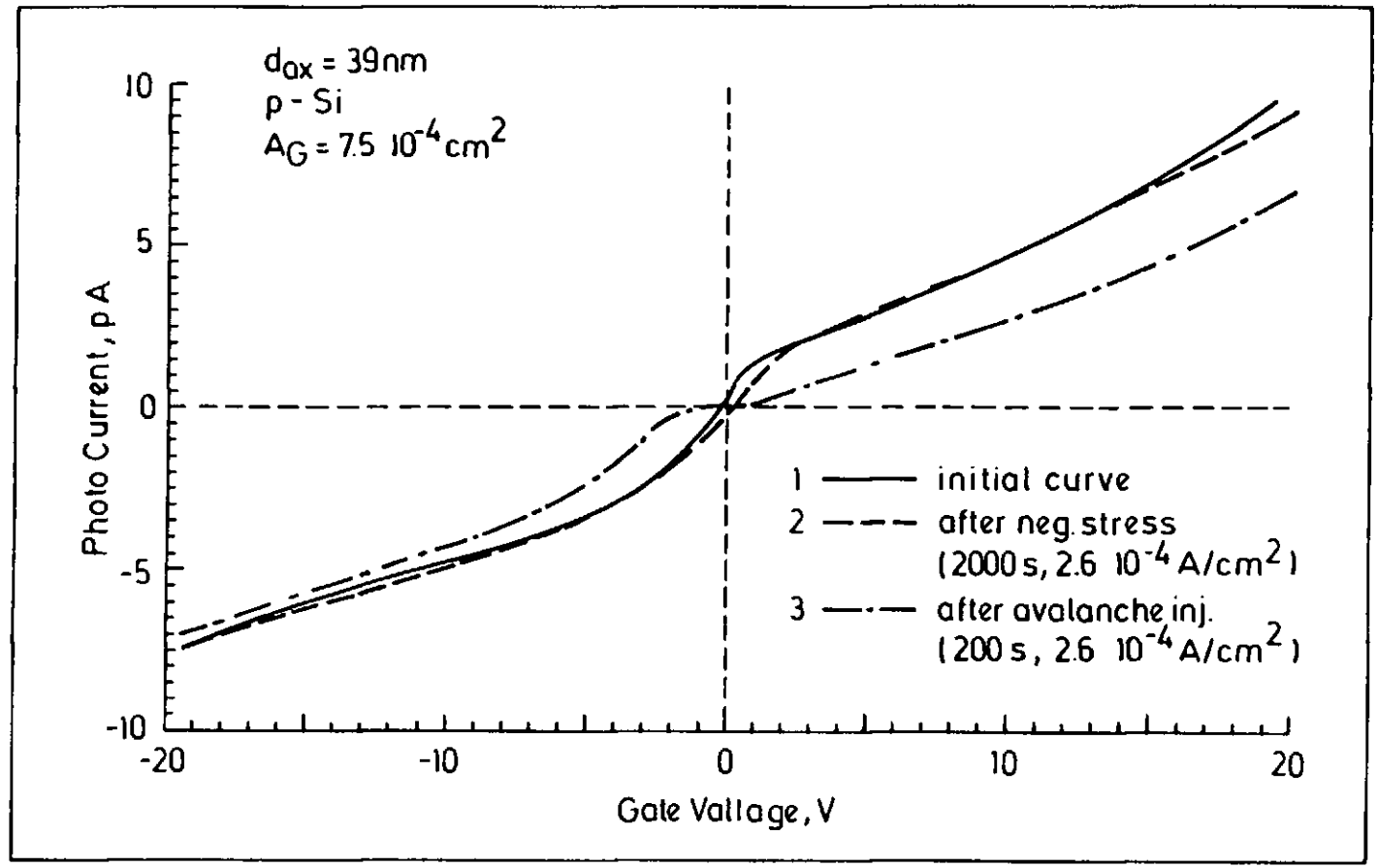


Figure 3.14 Photo I-V curves taken on wafer MP4 cap H (1) before stress (2) after negative constant current stress (3) after subsequent avalanche injection.

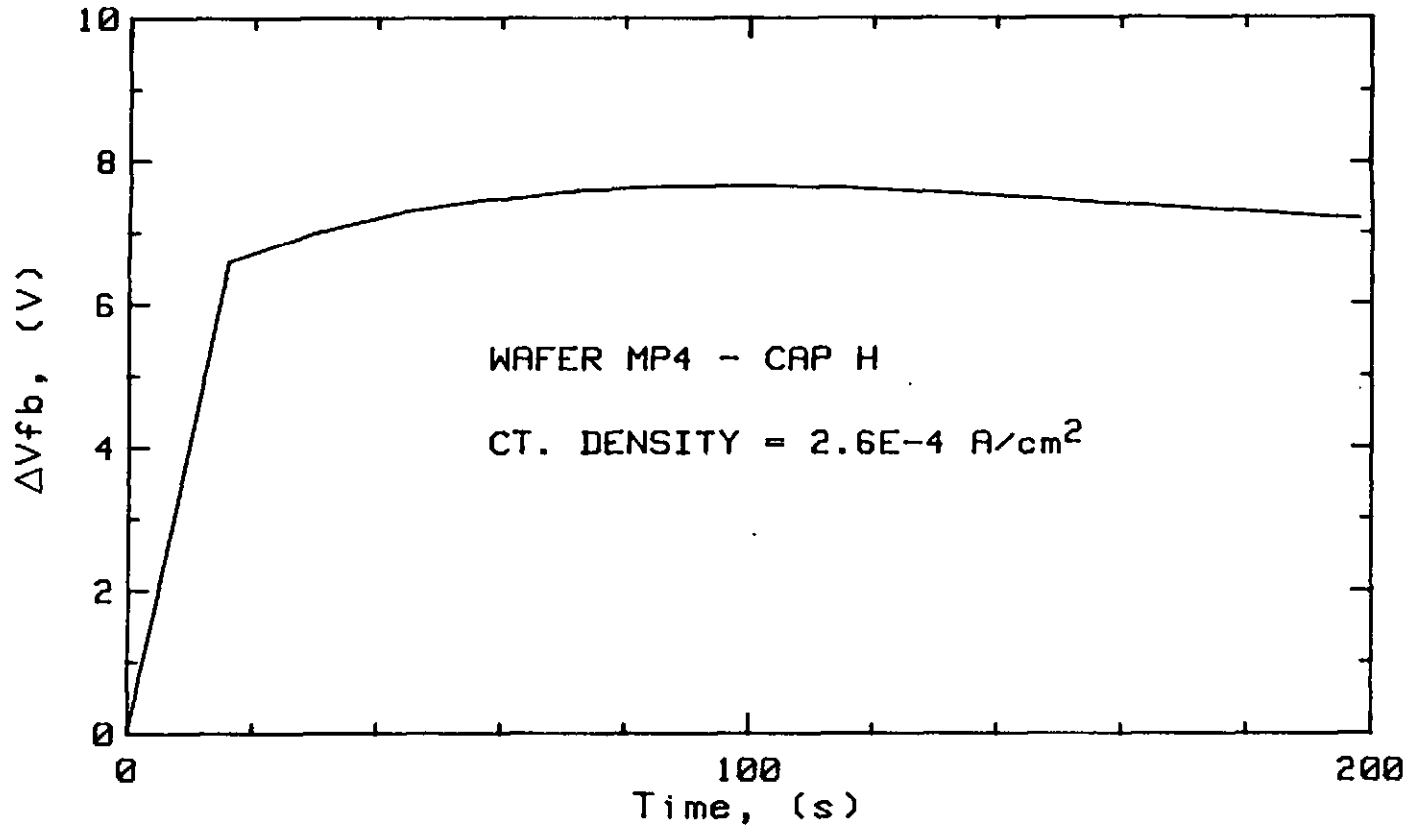


Figure 3.15 ΔV_{fb} as a function of time during avalanche elctron injection into cap.H, wafer MP4. Previously a negative voltage constant current stress had been applied at $2.65 \times 10^{-4} A/cm^2$ for 2000s.

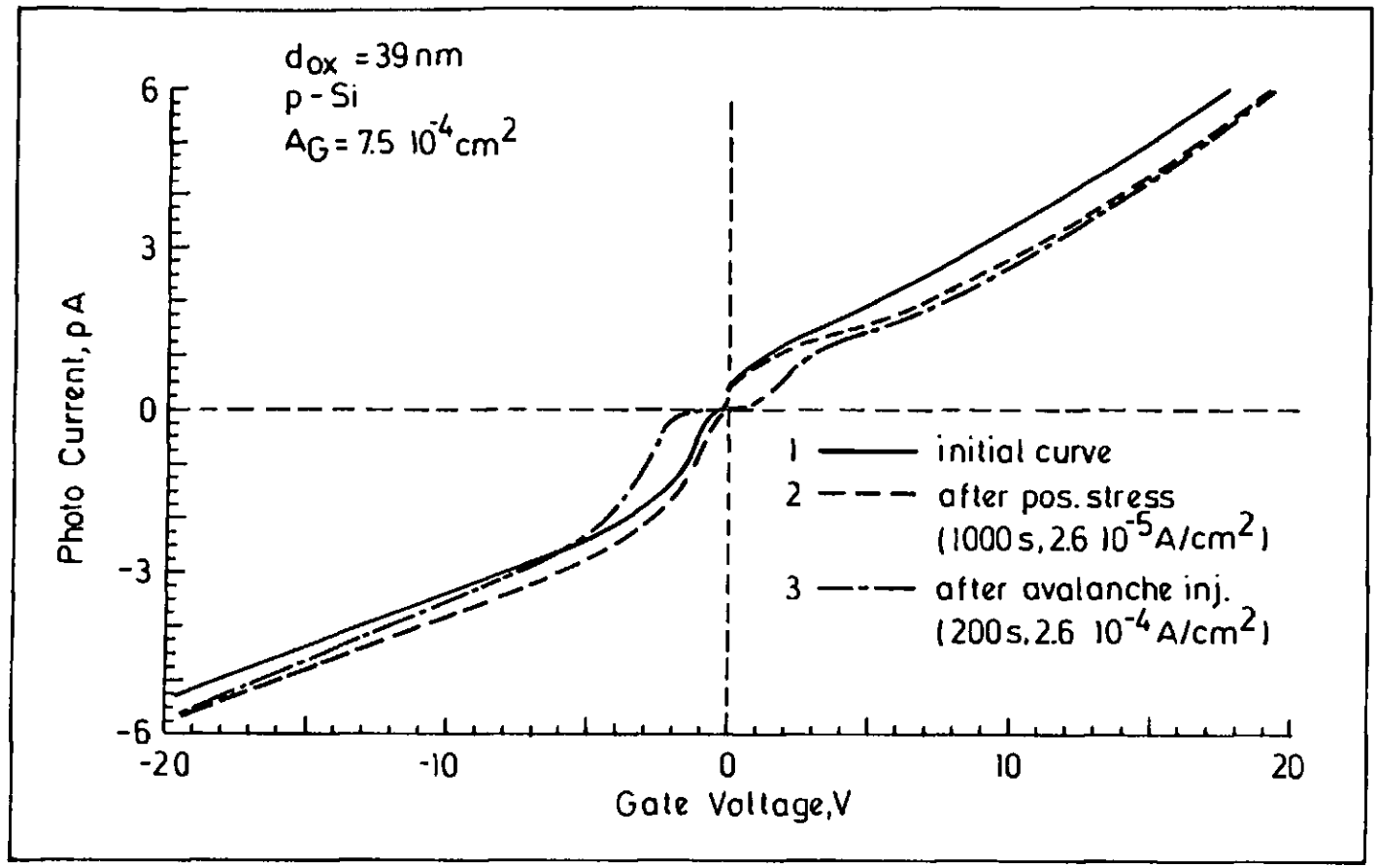


Figure 3.16 Photo I-V curves taken on cap.H, wafer MP4: (1) before stress (2) after positive constant current stress (3) after subsequent avalanche injection of electrons.

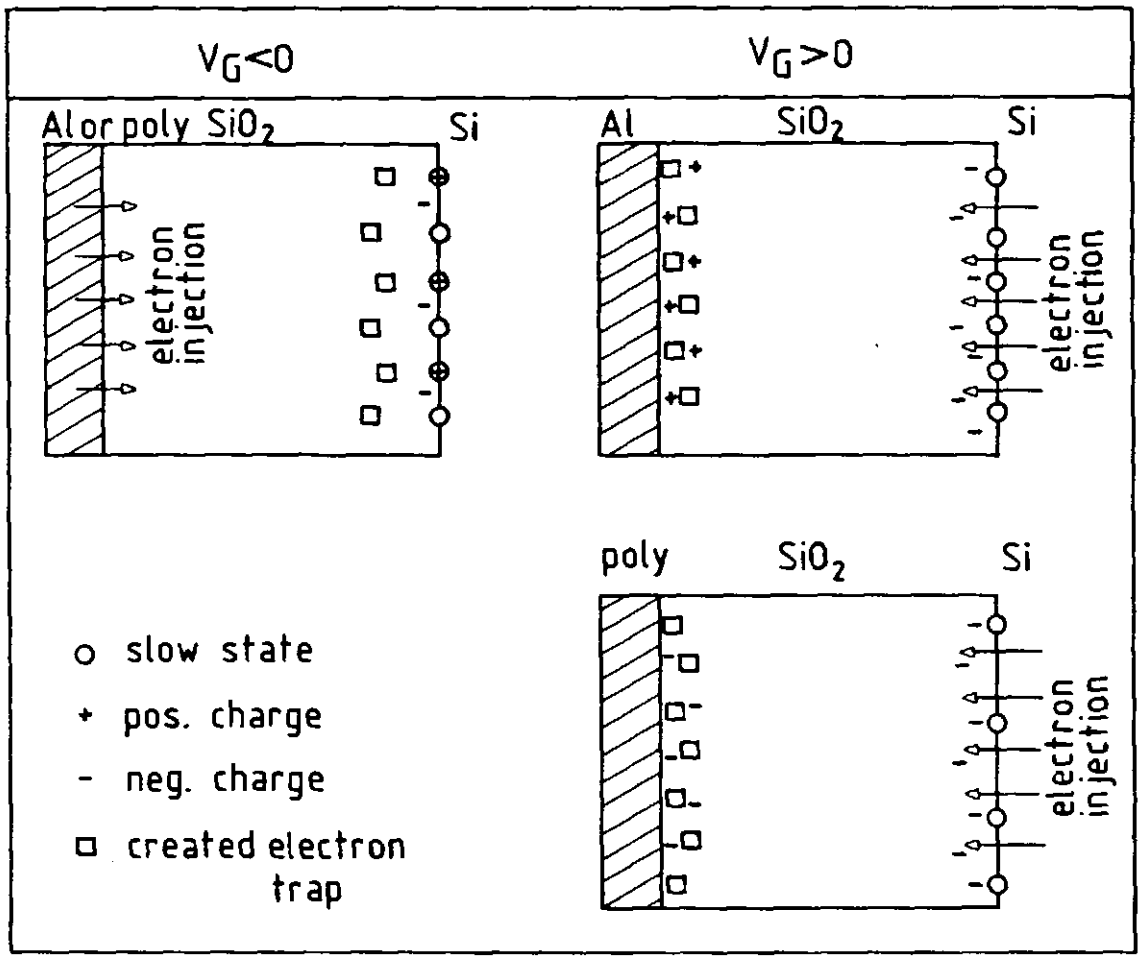


Figure 3.17 Schematic illustration of the charge and defect distribution after negative ($V_G<0$) and positive ($V_G>0$) constant current stressing of MOS capacitors with aluminium and polysilicon gate electrodes.

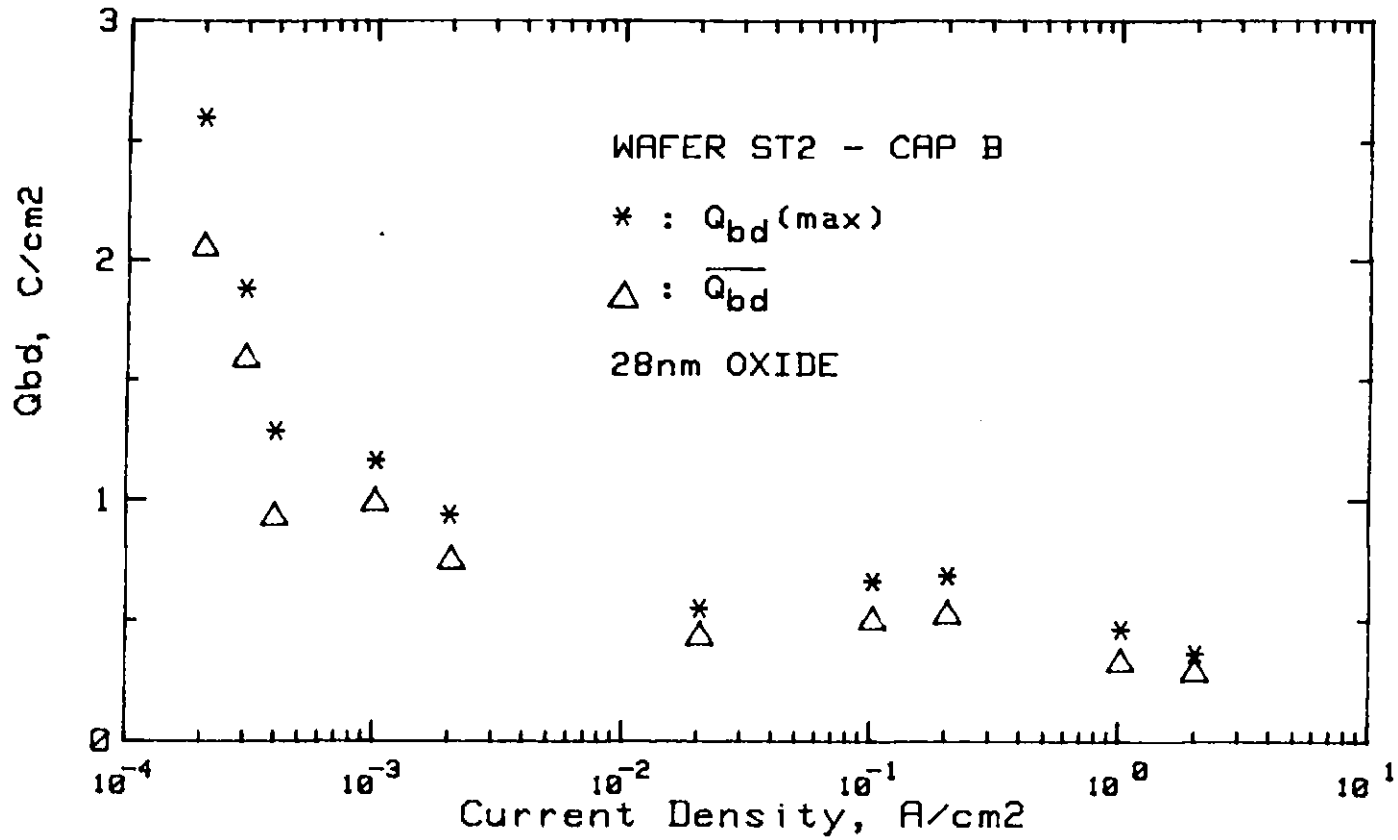


Figure 3.18 Maximum (Q_{bd}(max)) and average ($\overline{Q_{bd}}$) values of total charge to breakdown as a function of the current density - wafer ST2, cap.B.

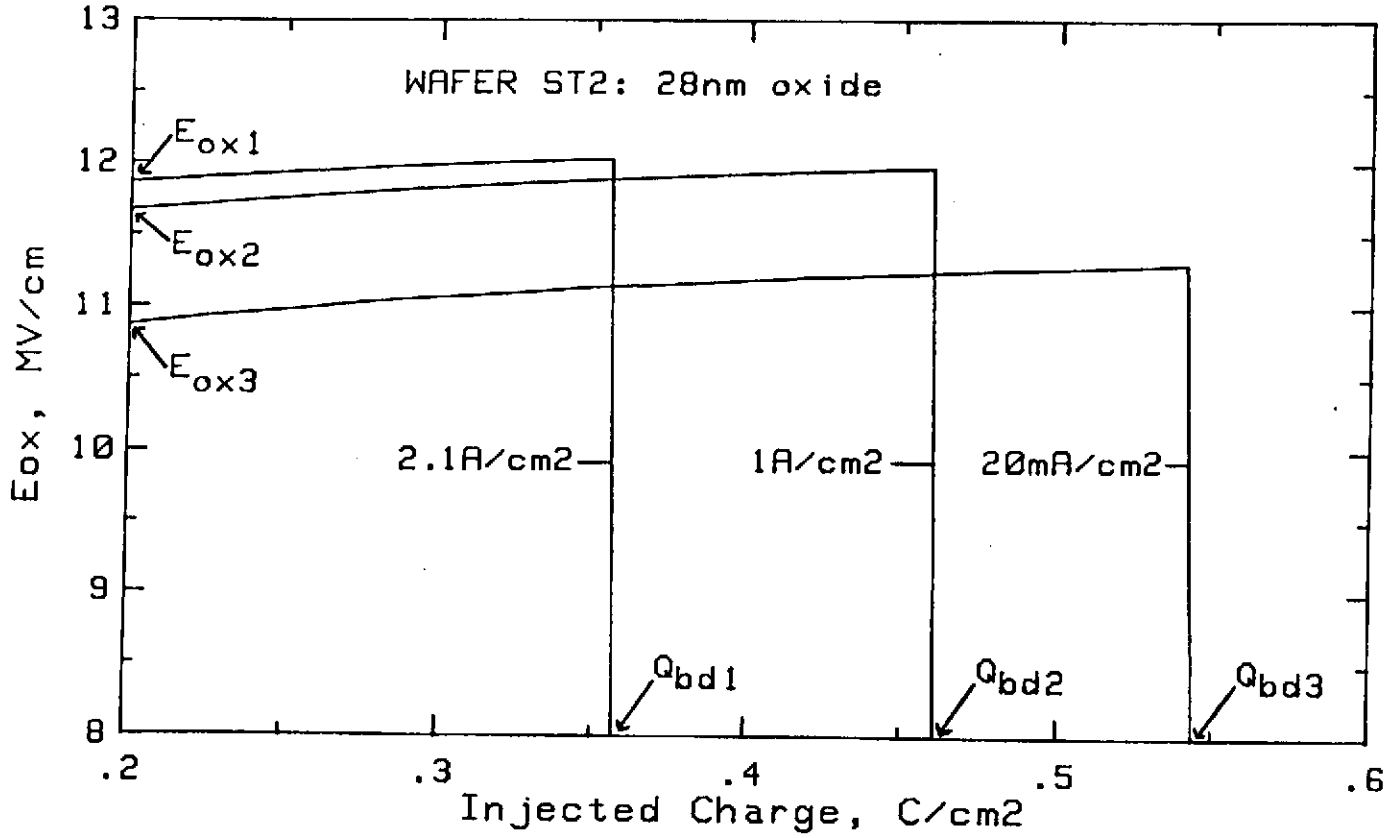


Figure 3.19 Extraction of E_{ox} and Q_{bd} from measurement of the applied field as a function of injected charge during high field constant current stress.

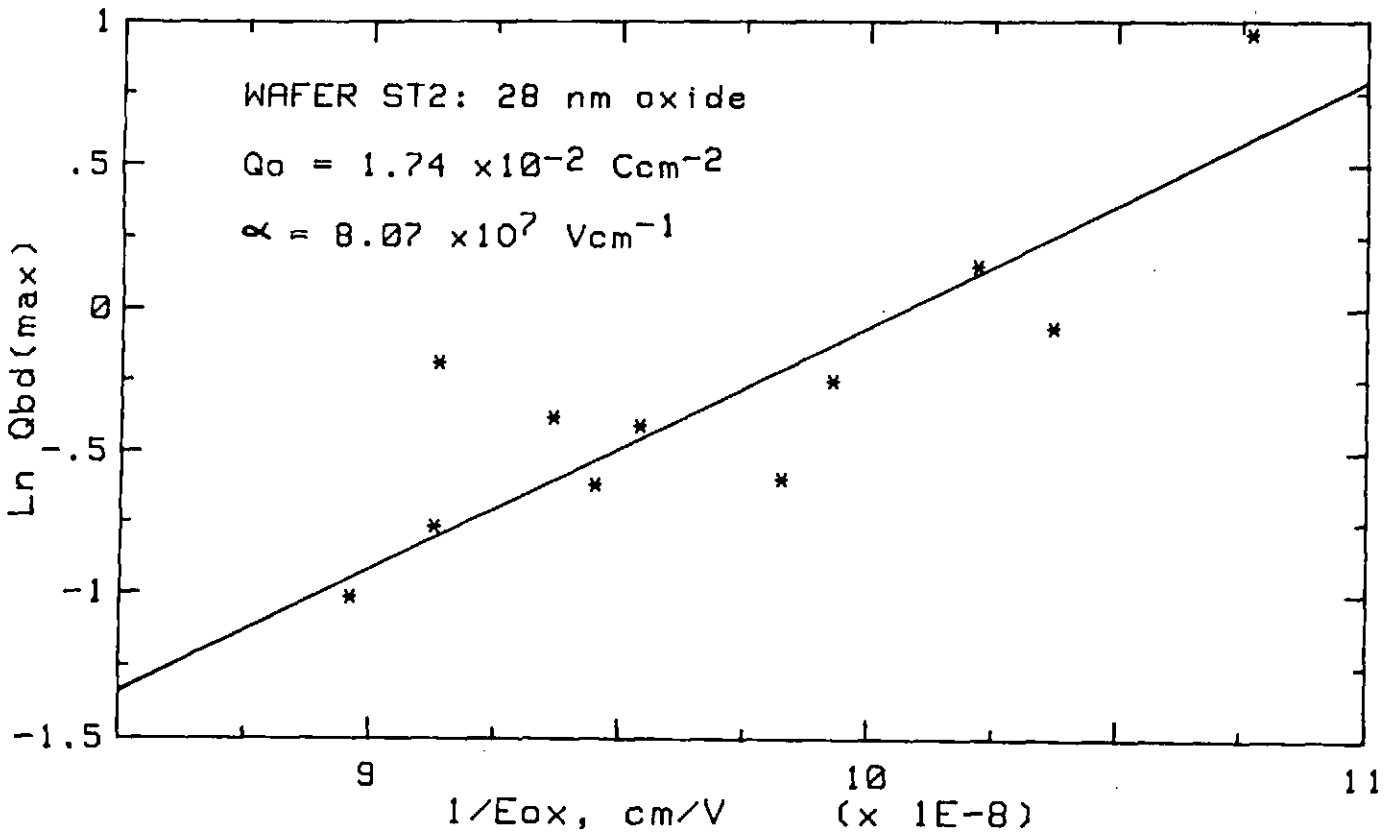


Figure 3.20 Ln (Q_{bd}(max)) as a function of $1/E_{ox}$. E_{ox} is varied by stressing at different current densities.

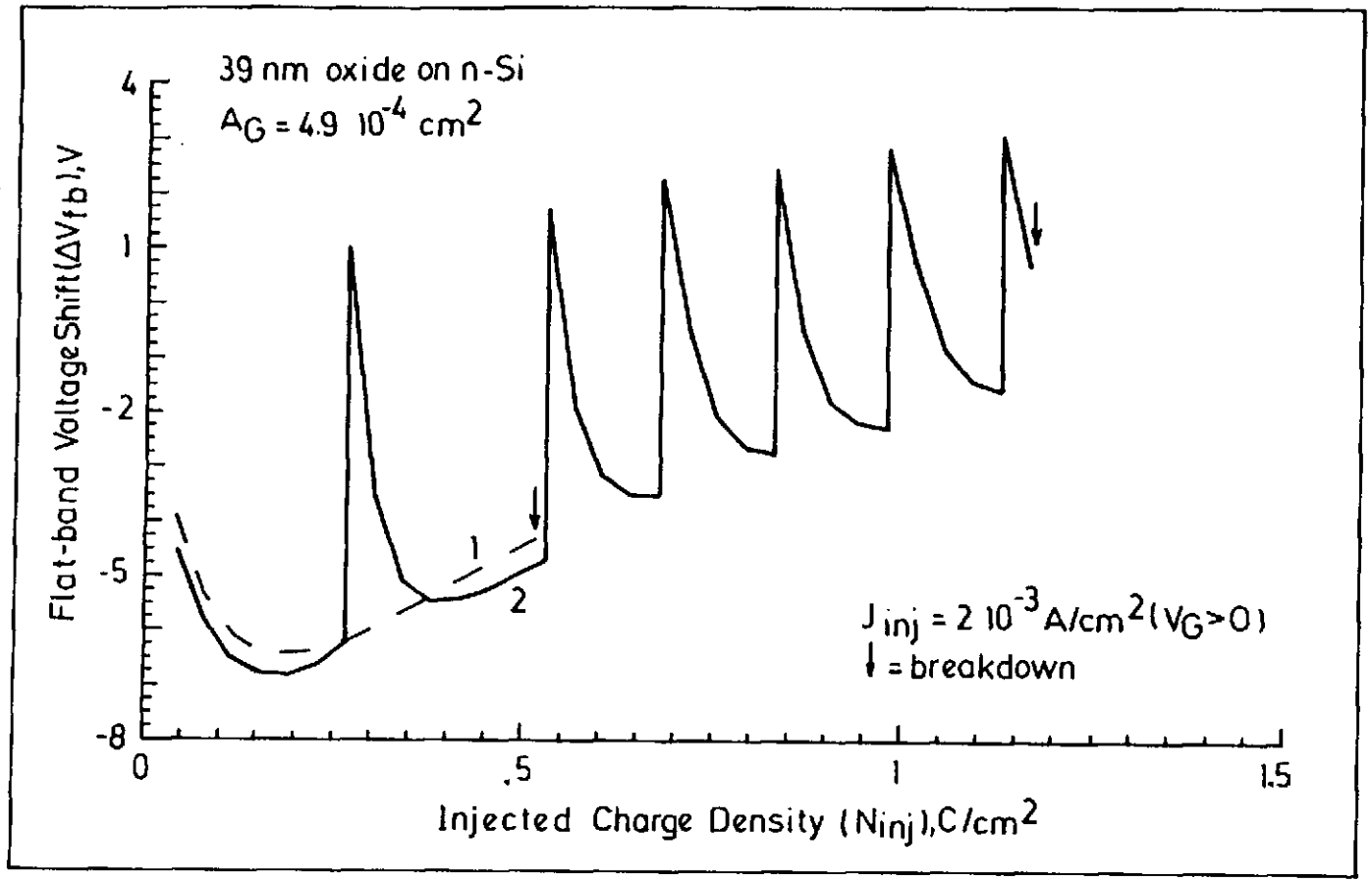


Figure 3.21 ΔV_{fb} as a function of injected charge, showing the point of breakdown for cap.B on wafer L9. (1) for continuous constant current stress and (2) or periodic interspersal of +5MV/cm during the constant current stress.

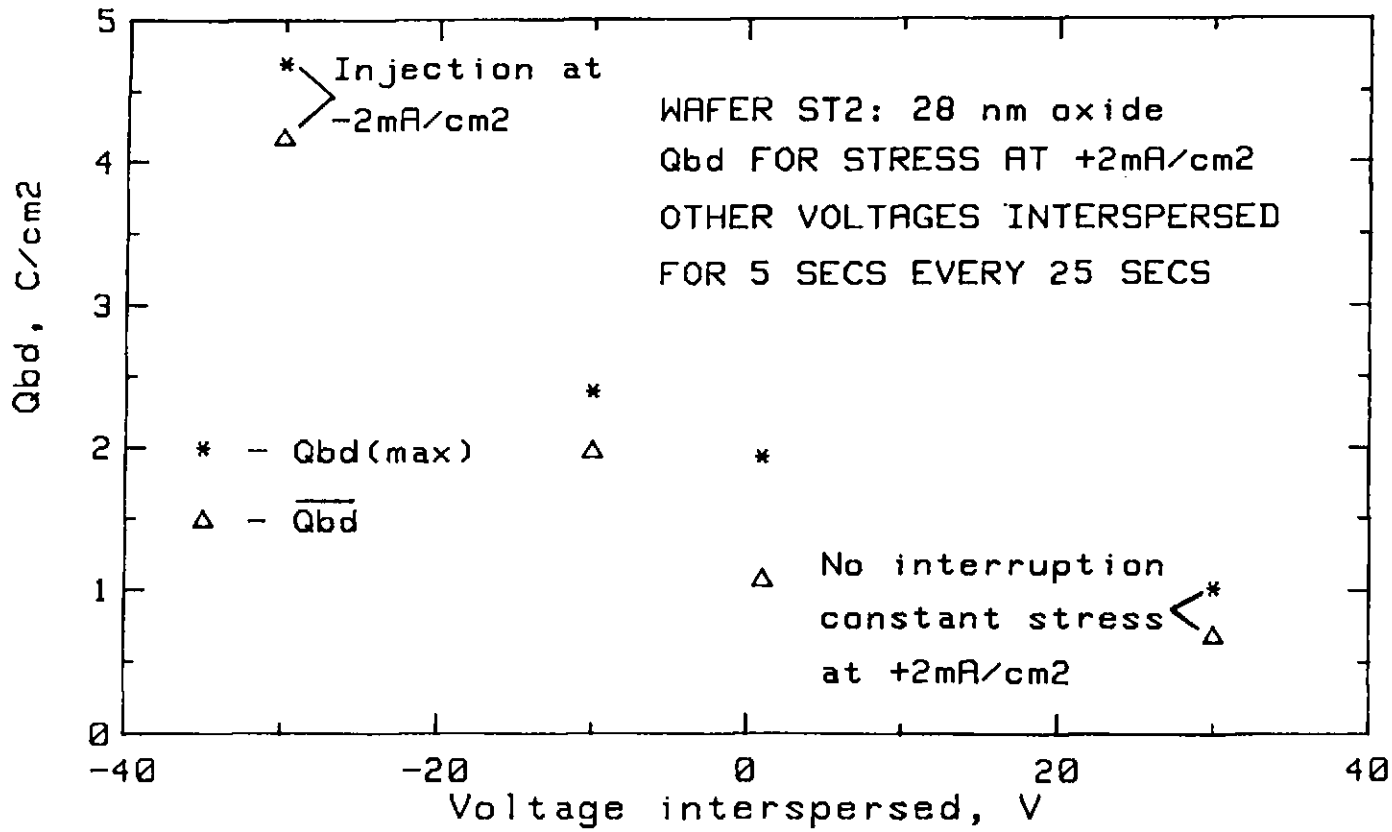


Figure 3.22 Maximum (Q_{bd}(max)) and average ($\overline{Q_{bd}}$) values of Q_{bd} at a stress of +2mA/cm^2 as a function of the voltage interspersed for 5s every 25s.

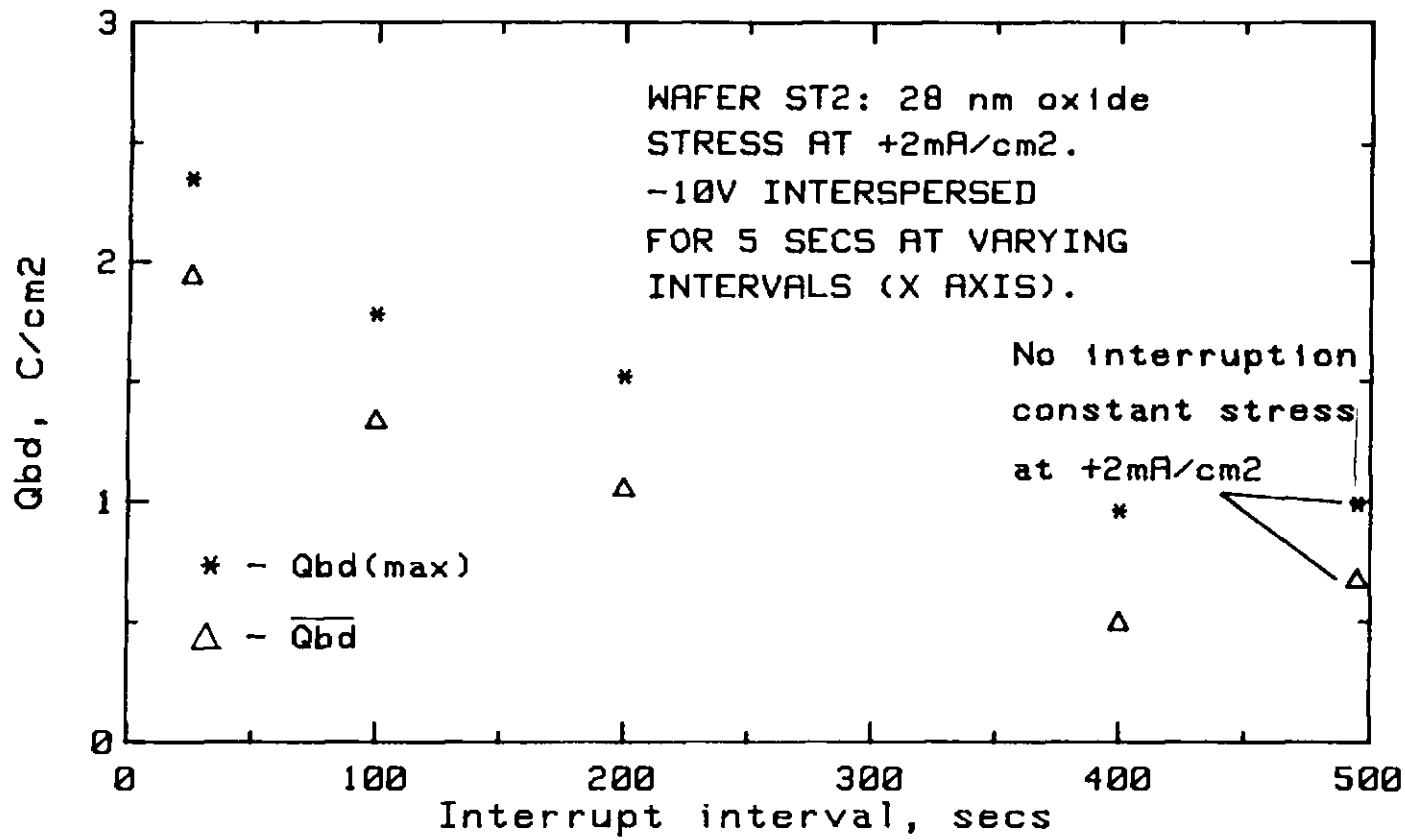


Figure 3.23 Maximum (Q_{bd}(max)) and average ($\overline{Q_{bd}}$) values of Q_{bd} at a stress of +2mA/cm^2 interspersing -10V for 5 secs at varying intervals.

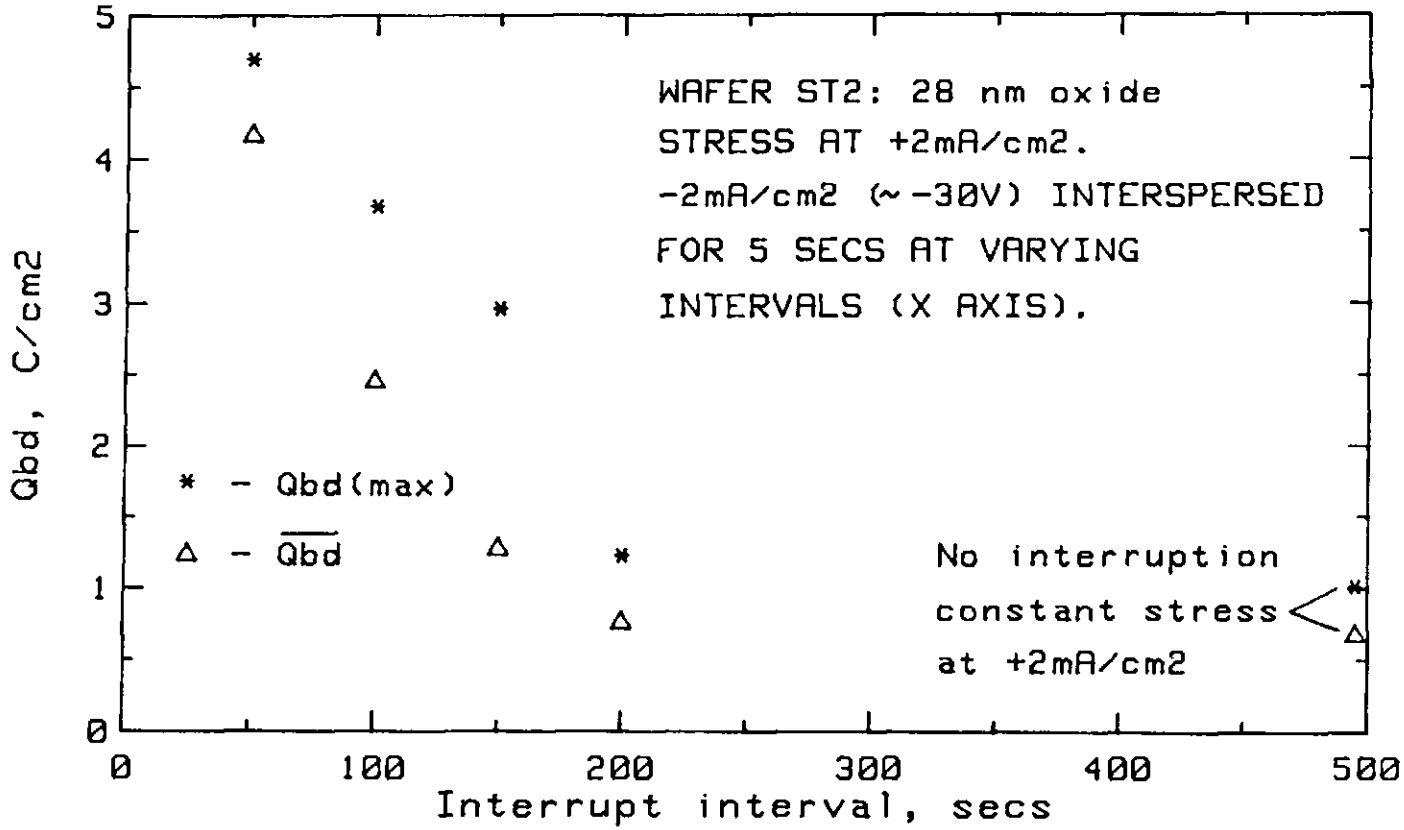
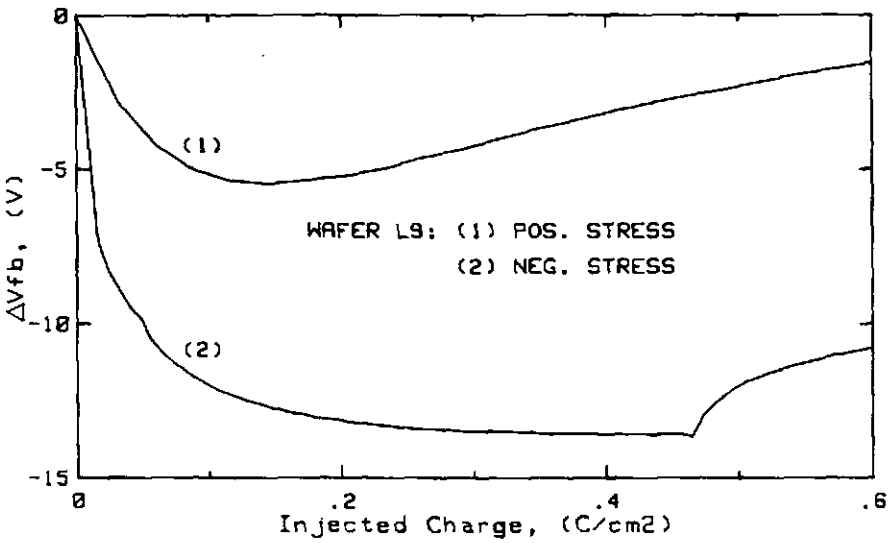


Figure 3.24 Maximum (Q_{bd}(max)) and average ($\overline{Q_{bd}}$) values of Q_{bd} at a stress of +2mA/cm^2 interspersing -2mA/cm^2 for 5 secs at varying intervals.

(a) ALUMINIUM GATE



(b) POLYSILICON GATE

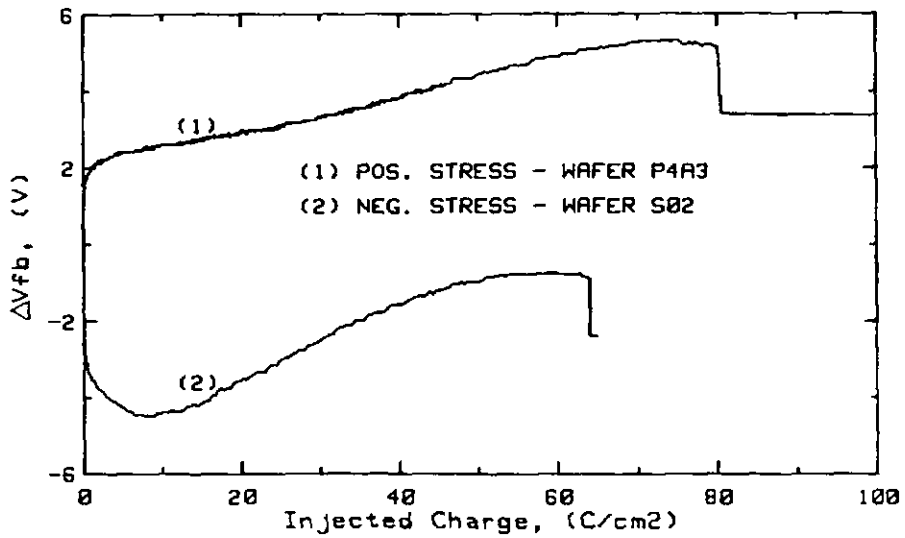


Figure 3.25 ΔV_{fb} as a function of injected charge for n-type capacitors stressed at $2 \times 10^{-3} A/cm^2$ at both gate polarities. (a) Aluminium (wafer L9) and (b) polysilicon (wafers P4A3 and S02) gate electrodes.

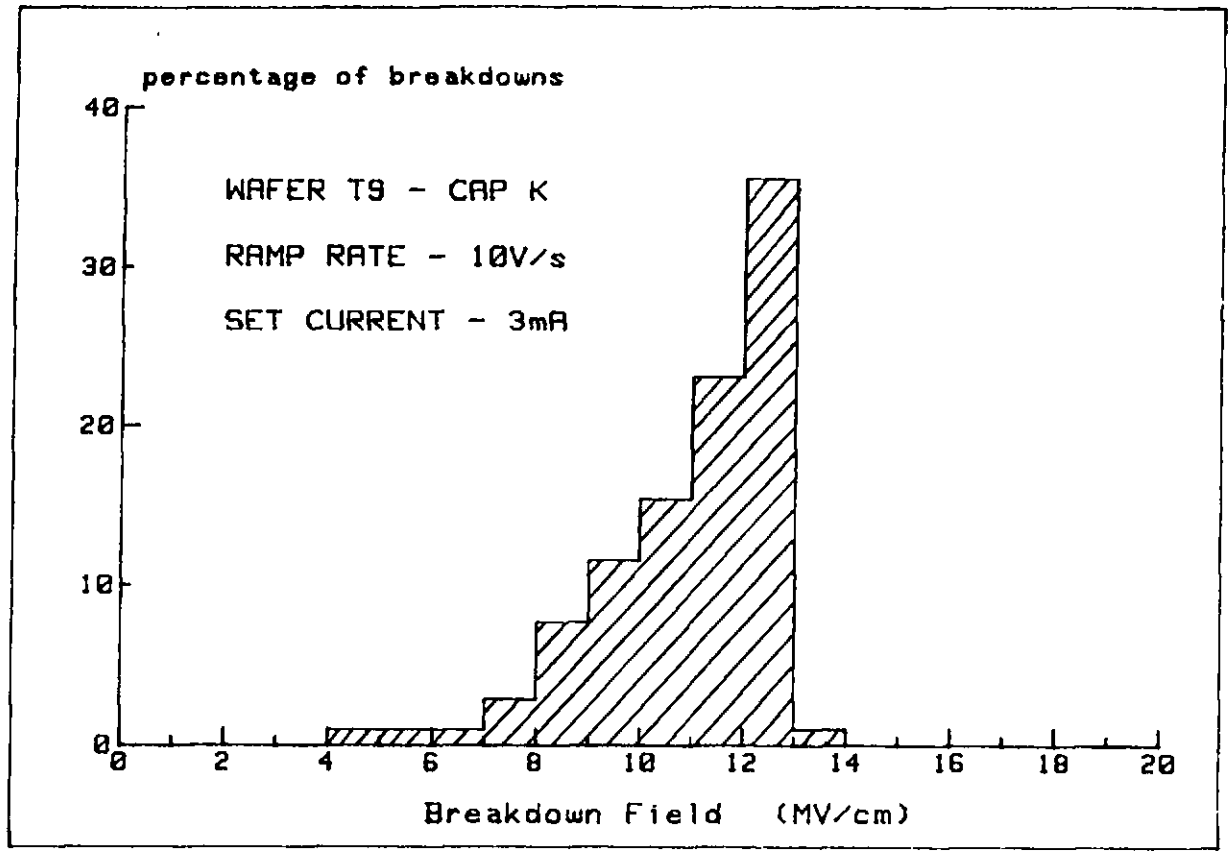


Figure 3.26 Breakdown histogram from a dielectric strength measurement on wafer T9, cap K, at a ramp-rate of 10V/s.

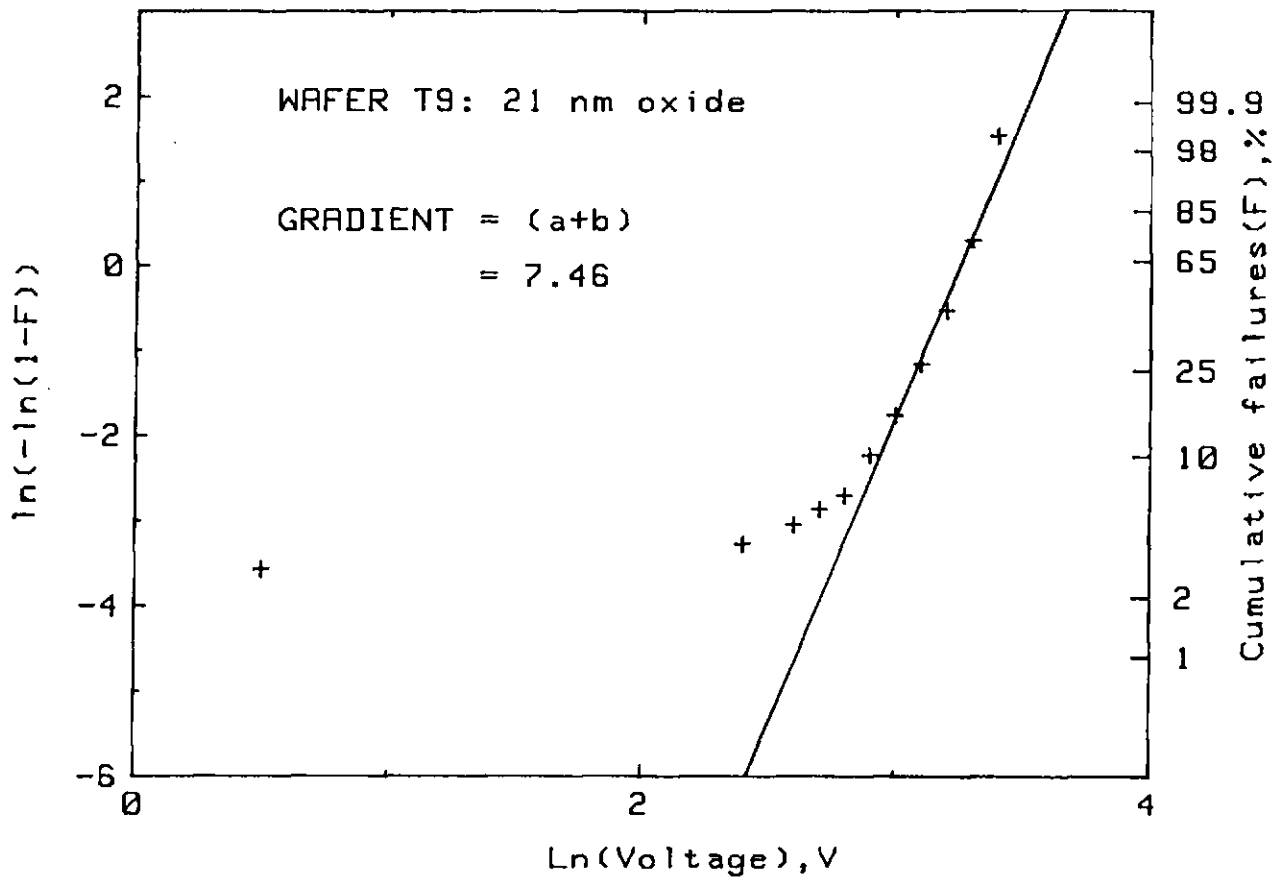


Figure 3.27 Weibull plot for a dielectric strength measurement on wafer T9, cap K, at a ramp-rate of 10V/s (data contained in breakdown histogram of Fig.3.26). F = cumulative probability of failure.

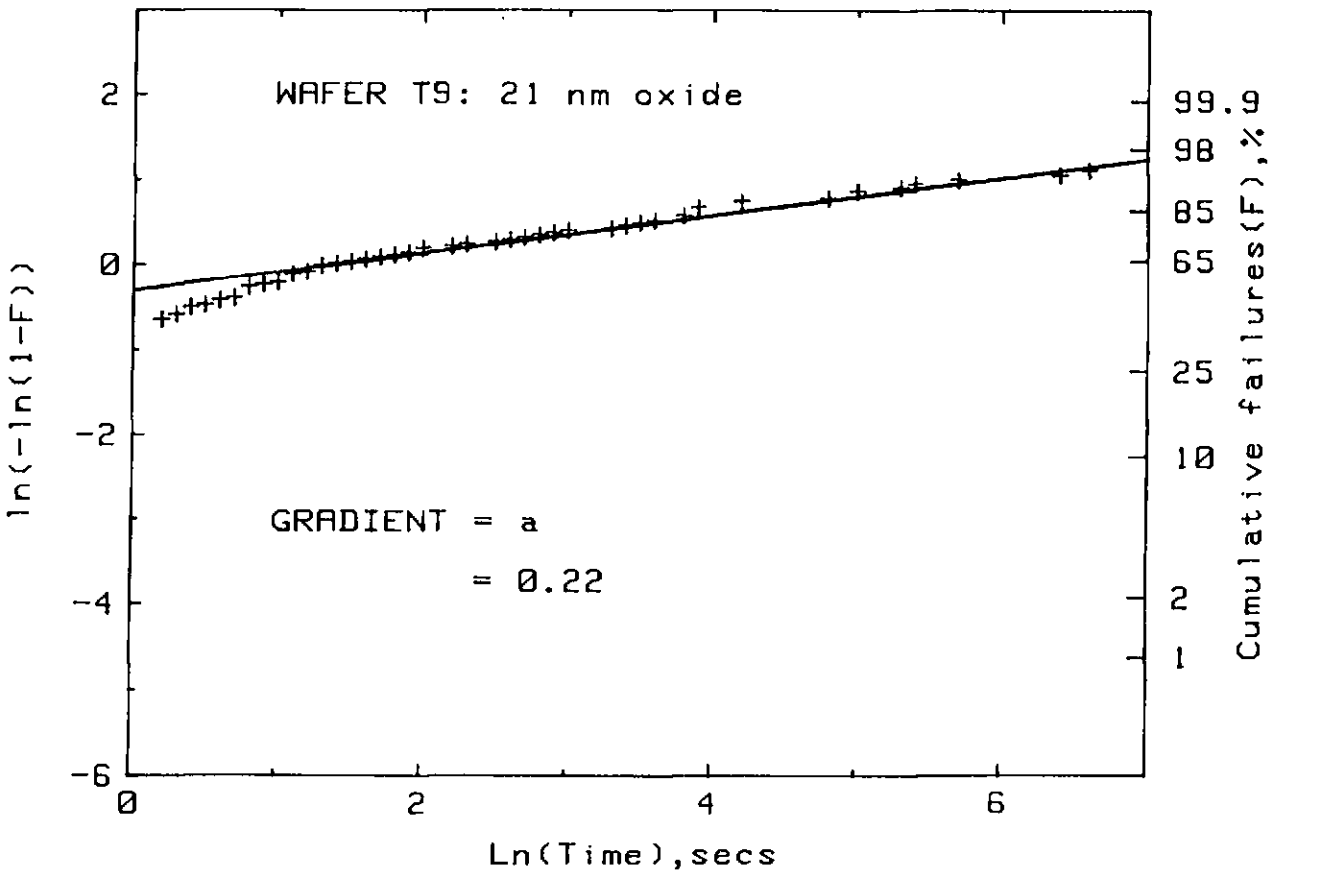


Figure 3.28 Weibull plot for a wearout measurement on wafer T9, cap.K at an applied field of 10.5MV/cm. F = cumulative probability of failure.

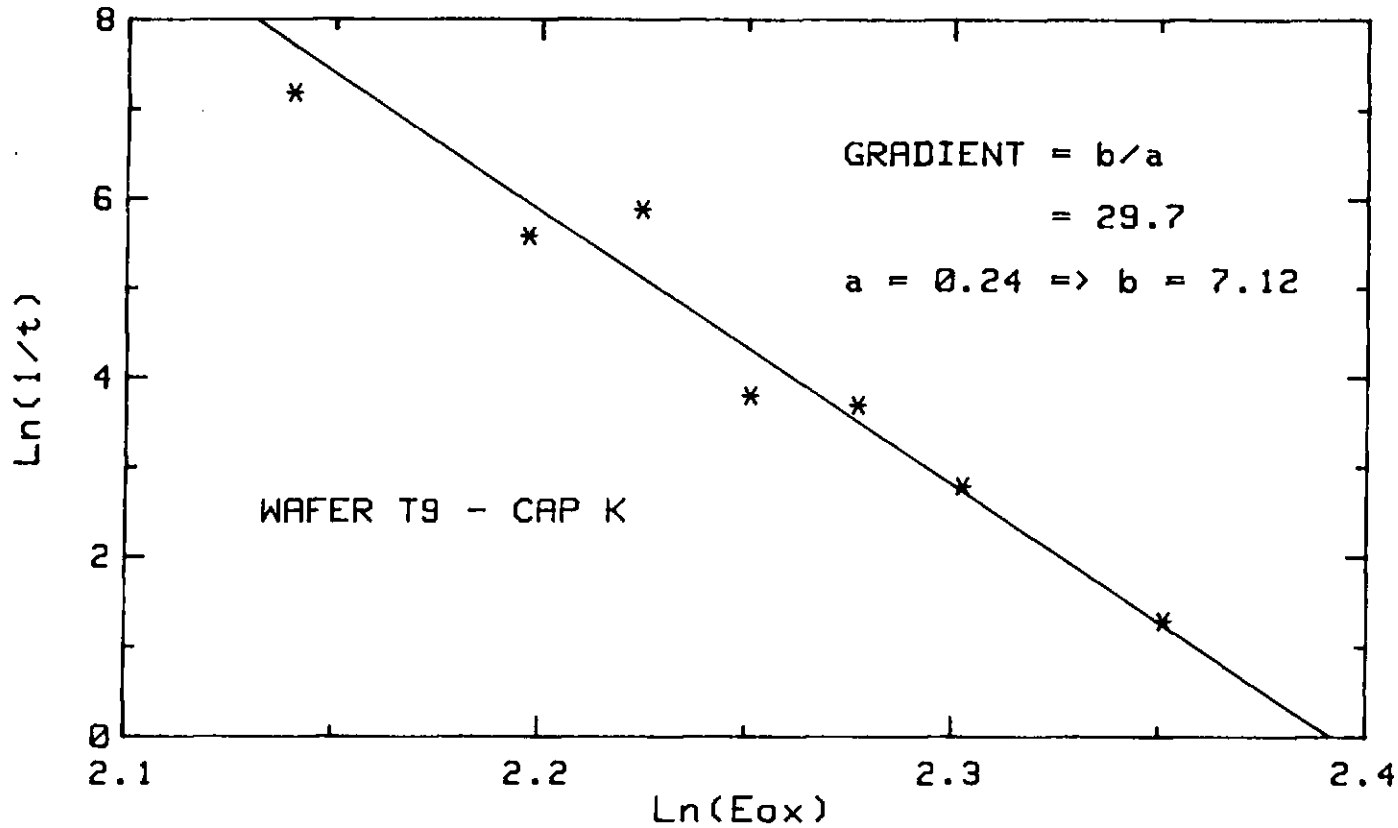


Figure 3.29 Plot of Ln(1/t) versus Ln(E_{ox}) for wearout measurements conducted at various applied fields on wafer T9, cap K.

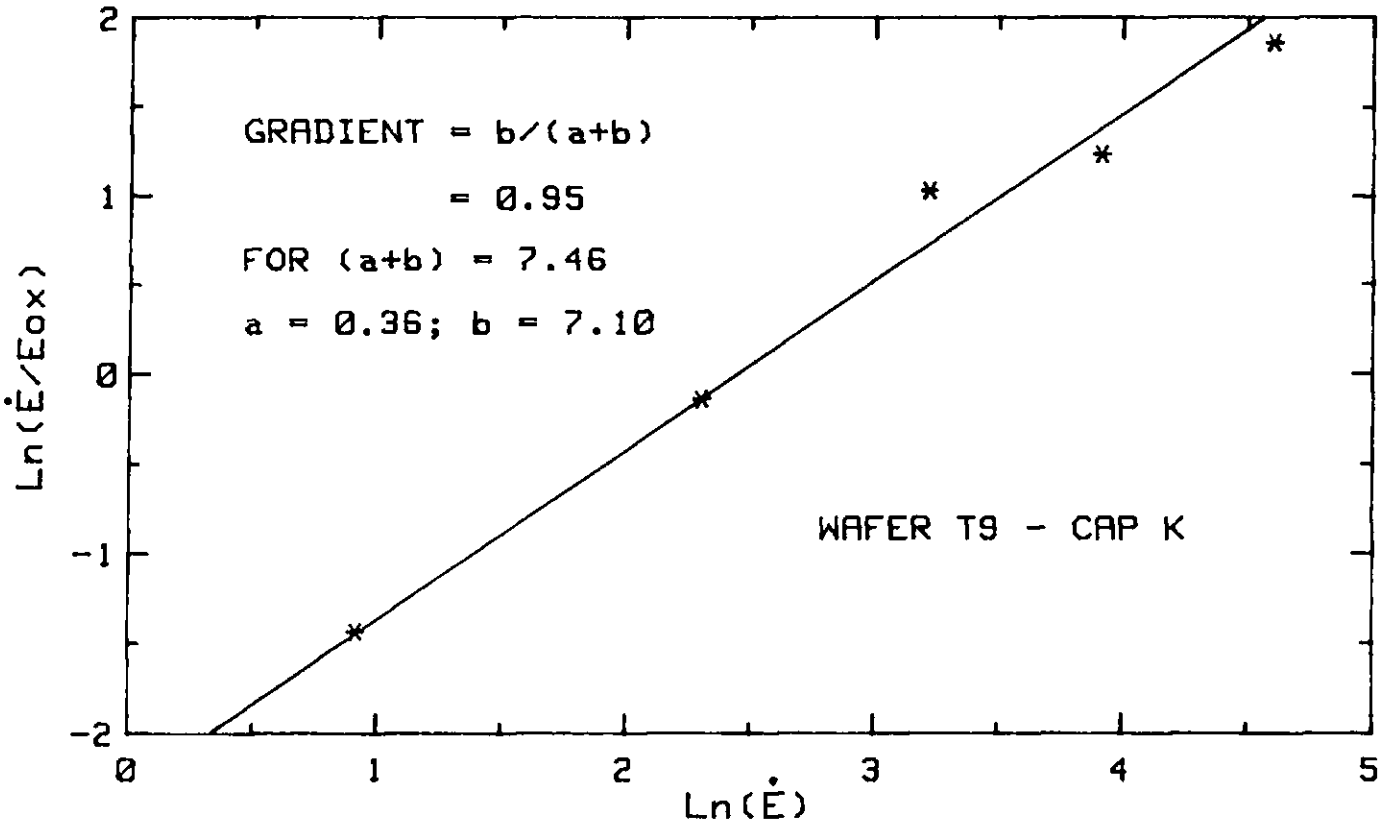


Figure 3.30 Ln $(\dot{E}/E_{ox})$ versus Ln $(\dot{E})$ for dielectric strength measurements conducted at different ramp-rates on wafer T9-cap K.

CHAPTER IV:
THE NEGATIVE BIAS INSTABILITY

4A: LITERATURE SURVEY

Application of a bias at elevated temperature is a simple and convenient method of achieving accelerated ageing of MOS capacitors. At high biases (usually >5 MV/cm) and room temperature or above, it can be used to examine the statistics of TDDB as discussed in the previous chapter. Using more moderate biases, measurement of the high frequency C-V curve at intervals allows the oxide charging kinetics to be studied. The former gives information largely concerning the expected lifetime of the device and the events immediately prior to breakdown whereas the latter indicates the degradation which can be expected to occur during the working life of the device and any potential threat this may offer to its normal operation.

Numerous studies have been made of the kinetics and processing dependence of both the shift in the C-V curve and of interface state generation occurring during both positive and negative BTS (48,49,58,62-67,167,168). Interface state density distributions over the silicon band gap are usually obtained using one of the methods described in chapter II, i.e. comparison of the HF and quasi-static C-V curves or of the ideal and quasi-static C-V curves. Since for positive bias the effects are small, unless there is contamination from mobile ions, emphasis will be placed here on what is known as the negative bias instability. Measurements made on transistors (60,61,169-171) will be largely ignored since the application of a source-drain bias allows the possibility of hot electron injection from the channel. This complicates the picture and makes it difficult to compare data from capacitors and transistors. In the next section (4A.1) the phenomenology and kinetics of the negative bias instability will be considered. Processing factors influencing both $Q_{ot}(+)$ and D_{it} generation are then discussed in section 4A.2 and models for the instability in 4A.3.

4A.1 Bias-temperature stressing of MOS capacitors.

4A.1.1 Measurement of the voltage shift.

There is some discrepancy in the literature in the parameters which have been chosen to monitor the negative bias instability and also in the equations which have been chosen to fit the acquired data. The most commonly used parameter has been the flatband voltage shift (ΔV_{fb}). However, this has the disadvantage that it includes components due to both the positive oxide charge ($Q_{ot}(+)$) and to the effect of interface states (D_{it}). The threshold voltage shift (ΔV_T) is the parameter of most interest from the point of view of device applications but this suffers from the same problem as ΔV_{fb}. Some authors attempt to separate the effects of $Q_{ot}(+)$ and D_{it} by using the midgap voltage shift (ΔV_{mg}) as being representative only of the positive oxide charge (48,49). This will be the case if all the interface states above midgap are acceptors and all those below midgap are donors (172-174) (or indeed vice versa providing that there are equal numbers of states above and below midgap). It is shown empirically in section 4B.1.2 using results from the present work that it is indeed a reasonable assumption that V_{mg} is a neutral point for interface states generated during BTS.

4A.1.2 Kinetics of ΔV_{fb} and ΔV_{mg}.

4A.1.2.1. Time dependence.

In terms of the mathematical form which the shift takes it appears not to be so important whether ΔV_{mg} or ΔV_{fb} is used since Haller et al. (48) found a similar time dependence ($\propto t^{0.2}$) of both these parameters i.e. the interface state density was increasing in proportion to the oxide charge generated. The kinetics of ΔV_{fb} and ΔV_{mg} will therefore be discussed together. $\Delta V_{fb} \propto t^{0.2}$ was also reported by Sinha and Smith (167) and Shiono et al. (63) found a similar relationship of $\Delta V_{fb} \propto t^n$ for their oxides grown in HCl,

where $0.31 > n > 0.2$ depending on the ageing temperature. For steam grown and dry oxides, the latter report a linear dependence of ΔV_{fb} on ln t. A similar relationship was found by Deal et al. (58), Hofstein (65) and Breed (66,67). Walden (175) has shown that such a logarithmic relationship would be expected if charge trapping in the oxide is responsible for ΔV_{fb} regardless of the exact form of the conduction law. The only requirements for this are that the current shows a strong dependence on the cathode field and hence decays as charge is trapped and that the charge trapping is permanent. The $t^{0.2}$ dependence is more difficult to account for by a theoretical model. A $t^{1/4}$ law would be compatible with a diffusion controlled process such as that suggested by Jeppson and Svensson for interface state generation (49) and the difference between $t^{0.2}$ and $t^{0.25}$ is certainly within the experimental error of the measurement. However, the difference between a ln t and a $t^{0.2}$ relationship is also not easy to distinguish experimentally since these are very similar functions.

4A.1.2.2 Temperature dependence.

The negative bias instability is obviously thermally activated and a good fit has usually been found to an Arrhenius law. However there is considerable variation in the values of the activation energy (E_a) obtained. Shiono et al. (63) found $E_a = 1.1$ eV for dry oxides and 1.5 eV for steam grown samples. Hofstein (65) also found $E_a \sim 1$eV while Sinha and Smith (167) report $E_a = 0.64$ eV. Breed (66) quotes two activation energies and postulates the existence of two centres in the oxide with ground states under flatband conditions below the top of the silicon valence band. Some of these centres are raised above the Fermi level at negative gate voltages. Transitions may then occur to excited states 0.6 eV (type 1) and 1.3 eV (type 2) above the ground state. Thermally assisted tunneling of electrons from the excited states to the interface is then responsible for positive charge generation. Breed conducted his measurements mainly at 77 K and he showed that considerable charge is generated at room temperature.

However, this is not observed unless the sample is cooled before C-V curves are measured since annealing also occurs rapidly at room temperature as soon as the bias is removed. The type 1 centre is unstable even at 77 K and discharges immediately when the gate and substrate are shorted.

The most recently reported activation energy is as low as 0.18 eV (59). This parameter obviously varies from batch to batch probably due to different processing conditions. Measurement procedure is also an important factor, however. For example, in reference (59) Haller et al. cool their samples with the gate floating thus allowing some time at elevated temperature for the charge to be annealed. Assuming that the rate of anneal increases with increasing temperature this could account for the extremely low value of the activation energy measured.

4A.1.2.3 Field dependence.

The field dependence reported for the negative bias instability also varies widely. The following relationships can be found in the literature: $\Delta V_{fb} \propto E$ (65); $\Delta V_{mg} \propto E^{3/2}$ (48) and $\Delta V_{fb} \propto E^{n}$ (167), where n is itself an inverse function of the ageing temperature. Between 100°C and 300°C n varied from 3 to 1.5. Again, the individual technology used seems to be an important factor.

4A.1.3 Generation of interface states during BTS.

Accompanying the positive charge generation under negative BTS is an increase in interface state density. Some increase is also seen under positive BTS conditions (63). Many authors have reported the generation of a peak in the interface state distribution at one or both polarities. Under negative BTS, Goetzberger et al. (64) found an interface state peak close to midgap. The exact position was dependent on processing details. In particular the electrode material was important but so also were the oxidation conditions. For

aluminium on steam grown oxides, the peak appeared at 0.3 eV above midgap while for aluminium on dry oxides it was at 0.25 ev below midgap. The remaining samples also showed peaks below midgap. Haller et al. (48) also report a peak below midgap, as do Shiono et al. (63) for all but their HCl grown oxides where there is a more uniform increase across the band gap. Jeppson and Svensson's dry oxides (49) developed a peak above midgap under negative BTS, at 0.65 eV from the valence band. The latter authors, along with Deal et al. (58), report not only a linear but a 1:1 relationship between midgap voltage and interface state density peak height.

Shiono et al. (63) report generation of a peak above midgap accompanying prolonged positive BTS. Haller et al. (48) also see a peak above midgap if first negative and then positive BTS is applied. Processing conditions are once more an important factor, playing a decisive role in determining the energy at which interface states are generated under BTS conditions.

4A.1.4 Kinetics of D_{it} generation.

Similar analysis has been applied to the kinetics of interface state generation under negative BTS as to the shift in V_{fb} and V_{mg}. A similar time dependence of the growth of interface states and of positive oxide charge is usually reported. Goetzberger et al. (64) found $\Delta D_{it} \propto \ln t$, while Haller et al. (48) found a $t^{0.2}$ dependence for ΔD_{it} as for $Q_{ot}(+)$. Shiono et al. (63) fitted their data to an exponential of the form $\Delta D_{it} \propto t^n$ where n is from 0.17 for steam oxides to 0.57 for oxides grown in HCl. Jeppson and Svensson (49) found a $t^{0.25}$ dependence at moderate to low fields and suggest a diffusion controlled reaction is responsible for the creation of interface states. This model is discussed further in 4A.3.2. Again, activation energies differ quite widely from each other. Goetzberger et al. (64) estimate a value of 1.4 - 1.5 eV while Jeppson and Svensson (49) find ~0.3 eV. Likewise there is considerable variation in the reported field dependence of interface

state generation. Goetzberger et al. (64) find $\Delta D_{it} \propto E$, Shiono et al. (63) $\Delta D_{it} \propto \exp(E^{0.5})$ and Haller et al. (48) $\Delta D_{it} \propto E^{3/2}$.

4A.2 Processing dependence of the negative bias instability.

$Q_{ot}(+)$ generated during negative BTS is roughly proportional to the fixed oxide charge, Q_f, initially present (58). Similarly, the number of interface states created, ΔD_{it}, is proportional to the initial interface state density, $D_{it}(init)$ (64). Hence, factors which affect Q_f and $D_{it}(init)$ (42) such as silicon orientation, oxidation ambient and post-oxidation anneal will also indirectly affect the negative bias instability. For example, Q_f is 2 or 3 times larger in oxides grown on <111> silicon than on <100> and correspondingly larger ΔV_{mg} values are observed during negative BTS on the <111> oxides. $D_{it}(init)$ is proportional to Q_f. Thus, $D_{it}(init)$ and ΔD_{it} during BTS also show this dependence on silicon orientation. Given the importance of Q_f and $D_{it}(init)$, the effect of processing parameters on them as well as on $Q_{ot}(+)$ and D_{it} will be considered.

4A.2.1 Oxidation conditions.

The well known Deal's triangle (Fig.4.1) describes the dependence of Q_f on the oxidation temperature and final ambient (42). Q_f is almost certainly due to some non-stoichiometry at the interface and the triangle can be explained in terms of the relative rates of diffusion of oxygen to the interface and its reaction with the silicon. At high temperatures, where oxygen diffusion is the rate determining step, most of the excess silicon reacts with oxygen leaving a minimum level of Q_f. At lower temperatures the silicon-oxygen reaction becomes rate-determining and excess silicon (a high Q_f) is built up at the interface. If the ambient is changed to an inert gas immediately after oxidation, the excess silicon reacts with the remaining oxygen at any temperature thus lowering Q_f.

D_{it}(init) shows a similar dependence on oxidation temperature. However, it reacts differently to a final anneal in nitrogen. If dry N_2 is used, D_{it}(init) tends to increase (42). This and the effect of other anneals will be considered further in the next section.

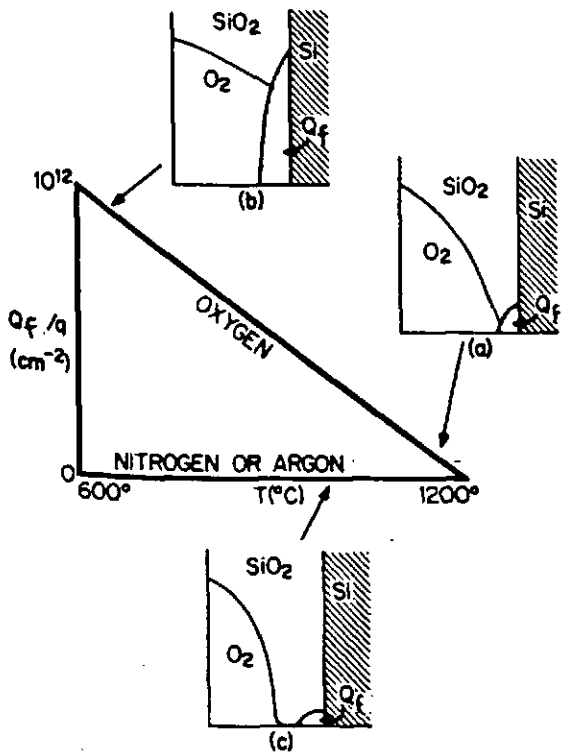


Figure 4.1 Deal's triangle, illustrating the dependence of Q_f on the oxidation temperature and final ambient (from (42)).

Water or HCl in the oxidation ambient are also important factors. The density of interface traps is strongly dependent on the partial pressure of water, being larger for dryer oxides (63). Breed and Kramer (62) report less variation of Q_f and Q_{ot}(+) with oxidation temperature for oxides grown in wet nitrogen than for those grown in dry oxygen (aluminium gate devices). Consequently, at high oxidation temperatures the density of oxide charge is larger in the wet oxides whereas at low oxidation temperatures the reverse is true. (Similar results have also been reported recently by Akinwande et al. (176).) For the wet oxides, the interface state density at midgap shows a

minimum at an oxidation temperature of about 900°C and the difference between <111> and <100> wafers is no longer significant.

For their polysilicon gate capacitors, Shiono et al. (63) also found a larger ΔV_{fb} during negative BTS for steam grown than for dry oxides. However, the dependence on stressing time was greater at higher temperatures for the steam oxides and at 300°C and 10^3 hours the shift was 3 or 4 times larger than for similarly stressed dry oxides. At the temperatures measured (200-300°C), the dependence of ΔD_{it} on stressing time was greater for the oxides grown in HCl than for the dry or steam grown oxides. Similarly, the HCl oxides showed a much more rapid increase in the rate of D_{it} generation with increasing applied field.

4A.2.2 Effect of anneals.

There are some apparently contradictory results in the literature on the effect of annealing on Q_f and D_{it}(init). The previous history of the sample as well as the precise time and temperature of the anneal all affect the result. While a short high temperature anneal in nitrogen reduces Q_f, an extended anneal causes it to start to increase again (42) (cf. effect on breakdown in 3A.1.2 and 3A.2.2). Montillo and Balk (177) found the effect of such an anneal on D_{it}(init) depended on the temperature used. Up to 700°C, D_{it}(init) increased and above this temperature it decreased. They also studied the effects of other high temperature anneals (>600°C) in oxygen, in helium and in a vacuum. Annealing in helium and in nitrogen gave similar results, as might be expected since both are rather inert gases. Vacuum annealing gave somewhat different results, however, causing a reduction in Q_f at either 600°C or 1000°C, whereas D_{it}(init) was increased. This was somewhat surprising since similar annealing behaviour would be expected in a vacuum as in an inert gas. However, the results in N_2 showed a good correlation with its water content, thus again highlighting the importance of this parameter and explaining the differing results obtained on annealing in apparently

similar ambients. In contrast to termination of the oxidation at the oxidation temperature an anneal in O_2 below the oxidation temperature always resulted in an increase in both Q_f and D_{it}(init).

Sinha et al. (168) examined the effect of a hydrogen anneal on Q_f, D_{it}(init) and on the negative bias instability. A high temperature anneal in H_2 (800-900°C; 1/2 hour) prior to aluminium metallisation significantly reduced Q_f and also ΔD_{it} during negative BTS. A low temperature anneal in H_2 tended to enhance both of these. D_{it}(init) was reduced by both anneals but following a low temperature anneal, more interface states were introduced by negative BTS. Breed and Kramer (62) also report that Q_f and $Q_{ot}(+)$ do not respond in the same way to a low temperature annealing step. (This time at 450°C in wet nitrogen.) Q_f is reduced by such a treatment whereas there is very little difference in $Q_{ot}(+)$ generated in those samples having received a low temperature anneal and in unannealed samples. They suggested that this difference may be because the states causing $Q_{ot}(+)$ are nearer to the interface than those causing Q_f.

4A.2.3 Gate electrode dependence.

Use of polysilicon as the gate electrode in place of aluminium greatly reduces the negative bias instability. The initial fixed oxide charge is likewise reduced by up to 40% (41). Doping of the polysilicon after deposition additionally reduces Q_f and greatly improves device stability (41). In the case of aluminium, geometry of the gate was also found to be significant in some cases. Breed and Kramer (62) report smaller values of D_{it} and Q_f in the samples with the largest gate area when the devices did not receive a low temperature anneal prior to encapsulation. They suggest lateral out-diffusion of some annealing species may be occurring during encapsulation. As has been previously mentioned the position of the peak generated during negative BTS has been found to be strongly dependent on the gate electrode (64).

4A.3 Models for the negative bias instability.

4A.3.1 Hole trapping model.

Hole trapping was first postulated by Hofstein in 1967 (65) as the cause of the negative bias instability. The model he proposed is illustrated in Fig.4.2. Donor states close to the $Si-SiO_2$ interface have an energy level close to the valence band edge. Since the potential well of the trap extends into the silicon a hole need not be excited into the oxide valence band to be captured. Instead communication with the silicon is by a mechanism similar to trap hopping. Under negative BTS conditions holes are excited over the trap barrier and are trapped in the oxide. Since the trap level lies near the Fermi level, the number of holes trapped and hence the flatband shift depends on the surface concentration and therefore on the applied bias. This model also explains the lack of dependence of the time constant on the applied field.

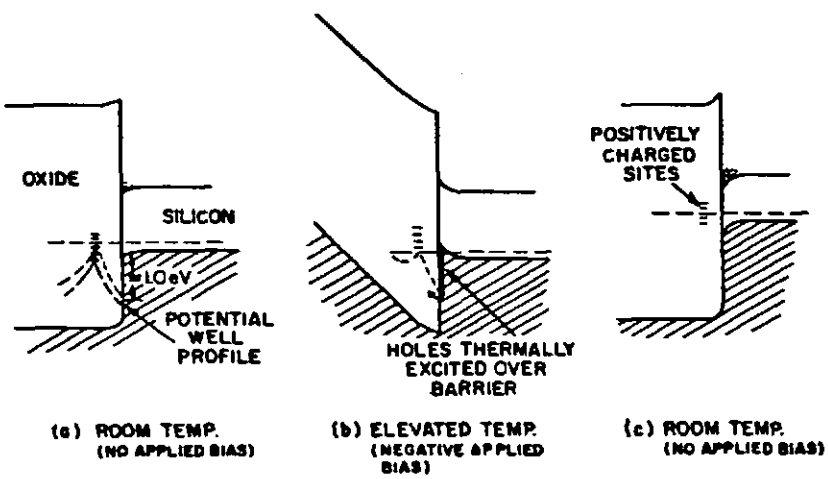


Figure 4.2 Hole trapping model for the negative bias instability (from (65)).

That the presence of holes in an oxide results in interface degradation is well established from other experiments. Hu and Johnson found a linear relationship between trapped holes and

interface states when holes were introduced into the oxide by irradiation or by high field stress (178,179). The generation rate was slow but after a year the ratio of holes to interface states reached 1:1. Lai also reports a linear increase in midgap interface state density with midgap voltage shift when samples are subjected to avalanche hole injection (47). Jeppson and Svensson (49) similarly found a 1:1 ratio between these two parameters after negative BTS which lends some support to the hole trapping model. However, they themselves suggest a hole tunneling process occurs only at high fields (>6 MV/cm) and low temperatures (-25°C). At lower fields and higher temperatures they suggest that a chemical reaction takes place resulting in electron emission. This is discussed further in the next section.

4A.3.3.2 Electron emission model.

A model based on electron emission was first described by Breed (66,67). As with the hole trapping model, the net result is the transfer of an electron from an oxide donor state to the silicon. The mechanism and kinetics of transfer are different, however. Breed's model is illustrated schematically in Fig.4.3 below.

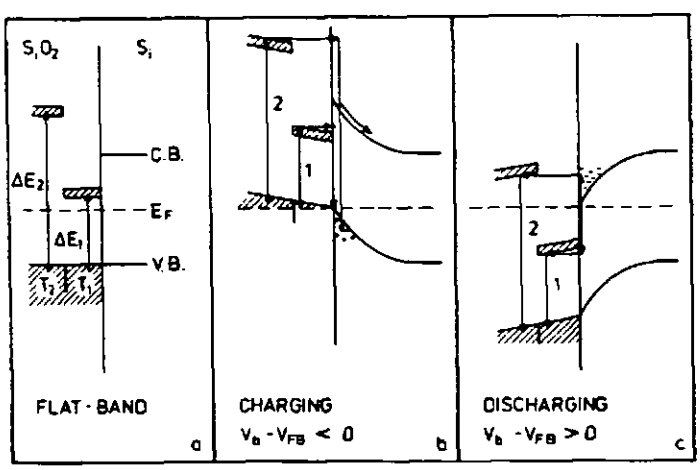


Figure 4.3 Electron emission model for the negative bias instability (from (66)).

Again, neutral centres are present near to the oxide-silicon interface which may act as donors. In the flatband condition (Fig.4.3a) the ground states of the traps are below the top of the silicon valence band and they have an excited state just above the silicon conduction band. Under negative BTS (Fig.4.3b) some of the ground states are raised above the Fermi level. Charging may then occur by a thermally assisted tunneling process involving thermal excitation of an electron to the excited state followed by tunneling of the electron to an interface state. At the interface either recombination may occur or the electron may be injected into the silicon. Population of the excited states is temperature dependent therefore, while the probability of electrons tunneling from the excited states to the silicon leads to a logarithm of time dependence.

The voltage shift due to positively charged centres is:

$$\Delta V_{fb}(t,T) = -(q/C_{ox}) \int N(x).dx \qquad (4.1)$$

where t is the time, T the temperature and N(x) is the density of active centres i.e. those whose ground states are raised above the Fermi level. For a homogeneous distribution of centres, $N(x) = pxV_a$, where p is a constant and V_a the applied bias. Under positive BTS (Fig.4.3c) all centres are below the Fermi level and there is a large concentration of free electrons at the surface. The charged centres are neutralised by tunneling of electrons from interface states to the excited states. For the type 2 centres, transition of an electron to an interface state from which it can tunnel is a thermally assisted process with an activation energy of about 0.4 eV.

The model of Jeppson and Svensson for interface state generation during negative BTS (49) has already been mentioned as an example of an electron emission process. At fairly low fields (<5 MV/cm) and elevated temperatures they suggest that the following chemical reaction is occurring:

$$\equiv Si_s\text{-}H + \equiv Si\text{-}O\text{-}Si\equiv \longrightarrow \equiv Si\cdot + \equiv Si^{+} + \equiv Si\text{-}OH + e^{-} \qquad (4.2)$$

Since the electron can be lost to the substrate only under negative bias there is a lack of symmetry of the interface state generation with respect to bias polarity. The $t^{1/4}$ dependence observed by these authors can be explained if the diffusion of Si-OH away from the interface is the rate determining step.

4A.3.3. Excess silicon model.

One of the models best abie to account for the observations on fixed oxide charge and the generation of further positive charge during negative BTS is the excess silicon model of Deal et al. (58). Thermal oxidation proceeds by diffusion of oxygen to the interface. Here there is excess silicon present for it to react with. When the oxidation stops, these excess silicon ions are 'frozen in', giving rise to positive charges (see Deal's triangle, section 4A.2.1). Application of a negative field at elevated temperature may break a second already strained bond to Si^{+}, resulting in an Si^{2+} species. This would explain the dependence of positive charge generation on initial levels of fixed oxide charge. In this model, interface states have a similar origin to Q_f, being associated with unsaturated silicon bonds at the interface. The role of a hydrogen anneal in reducing the density of these states can be easily explained, therefore.

Yamasaki studied the orientation dependence of the initial film growth rate, the shear modulus of a silicon substrate and their relationship to oxide charges (180). His results support the assumption that Si^{+1}, Si^{+2} and Si^{+3} exist at the interface. These species form the basis of the excess silicon model. More recently, the existence of the 1, 2, and 3 oxidation states of silicon at the interface has been demonstrated by XPS (181-183).

4A.3.4 Oxygen vacancy model.

Excess silicon and deficient oxygen appear to be two ways of describing the same situation. However, the energy levels of the two species are not necessarily the same. Bennett and Roth (68) used molecular orbital theory to calculate the energy of various defects in SiO_2, including an oxygen vacancy. The latter was found to be energetically quite favourable requiring three times less energy than formation of a silicon vacancy. The removal of an oxygen atom causes two energy levels to appear in the band gap, the upper level being doubly degenerate. If the lower level is unoccupied the neighbouring silicon atoms tend to move apart. The energy levels move upwards and the upper one merges with the continuum. In this case the empty level is above the silicon bandgap and would thus remain charged. This level could be responsible for fixed oxide charges.

The negative bias instability can be accounted for by the oxygen vacancy model in the following way. At sufficiently high temperatures and under the influence of an electric field, some oxygen vacancies are known to migrate (184). Under application of a negative bias, positive charge in the form of oxygen vacancies (V_o^+) could be generated by the following reaction:

$$xSi + SiO_2 \longrightarrow (1 + x)SiO_2 + 2xV_o^+ \qquad (4.3)$$

These vacancies then migrate slowly into the oxide. Under positive bias the vacancies drift back towards the $Si-SiO_2$ interface where they are eventually neutralised.

This model can account for all the non-oxygen annealing behaviour of Q_f; for example, the work of Fowkes and Hess on oxidation-reduction treatments (185). The smaller density of Q_f after wet oxidation can be attributed to hydrogen atoms bonding with the oxygen vacancies, while the dependence of Q_f on oxidation temperature can be related to the effect of stress on the vacancy.

The annealing experiments of Aslam and Balk (186) have provided experimental evidence for a relationship between oxygen deficiency and hole traps. A high temperature POA in O_2 was strongly correlated with high hole trap density. Conversely, a short final POA in O_2 substantially reduced hole trapping. $O_3{\equiv}Si\cdot$ $\cdot Si{\equiv}O_3$ and $O_3{\equiv}Si\text{-}O\cdot$ $\cdot Si{\equiv}O_3$ were postulated as likely hole trapping sites generated by the removal of either O_2 or water from the lattice. These sites are probably amphoteric, that is able to act as either electron or hole traps. This is compatible with the observed correlation between the density of deep hole traps and shallow electron traps following a given annealing treatment (186).

4A.3.5 Valence alternation pair model.

A model for defects in the $Si\text{-}SiO_2$ system based on valence alternation pairs (VAPs) has been proposed by Hubner (187). He attempts also to explain the generation of interface states and positive charge by irradiation and BTS using this theory. A VAP consists of a pair of oxygen atoms one of which is undercoordinated and one of which is overcoordinated, as illustrated below:

$$\begin{array}{l} {\equiv}Si \;.\;.\;. \; O^{-}\text{-}Si{\equiv} \\ \quad | \\ {\equiv}Si\text{-}O^{+}\text{-}Si{\equiv} \end{array}$$

These species ${\equiv}Si\text{-}O^{-}$ and ${\equiv}Si_3\text{-}O^{+}$ may act as hole and electron traps respectively. Interface states arise in this model from the dependence of the strength of the third $Si\text{-}O^{+}$ bond on its proximity to the interface. Due to increased dielectric screening and bond distortion at the silicon surface this bond is ten times weaker here than in the bulk, thus favouring ${\equiv}Si\cdot$ formation. The increase in D_{it} under BTS is explained as follows. Under negative BTS electrons captured by VAP defects are transferred to the silicon. An associated relaxation of ${\equiv}Si_3\text{-}O^{+}$ centres out of the interface plane releases a silicon dangling bond causing generation of an interface state at about 0.35 eV above the silicon valence band. Under positive BTS

electrons are transferred from the silicon to the weak $Si-O^+$ bonds generating an interface state at 0.42 eV below the silicon conduction band edge.

Fixed positive charge is attributed to uncompensated $\equiv Si3-O^+$ centres. The generation of positive charge by negative BTS is then accounted for by the breaking of weak Si-H bonds (presumably accompanied by loss of an electron to the substrate) and transfer of H^+ away from the interface. This increases the concentration of undercoordinated silicon and hence overcoordiated oxygen near to the interface. It is debatable however whether it is necessary to invoke hydrogen to account for $Q_{ot}(+)$ since loss of an electron to the silicon accompanied by rearrangement of the VAP can lead to generation of $\equiv Si-O^+$ and $\cdot Si\equiv$, thus accounting for both the positive charge and interface state generation during BTS.

On the basis of X-ray photo-electron spectroscopy (XPS) measurements Grunthaner et al. (143) also proposed the generation of these species at the interface on irradiation of the sample, a process which also leads to hole trapping and interface state creation. They suggest bond breaking is occurring in strained Si-O-Si bonds at the interface. This occurs via hole capture in the oxygen 2p lone pair orbital giving $O_3\equiv Si\cdot$ and $O_3\equiv Si-O^+$. The surrounding lattice undergoes a slight relaxation preventing immediate reformation of the bond. Subsequently the hole in the lone pair may be annihilated by an electron tunneling from the silicon. The resulting site is then $O_3\equiv Si\cdot \quad \cdot O-Si\equiv O_3$, which is one of the sites suggested by Aslam and Balk as a deep hole/shallow electron trap. Capture of an electron this site leads to $O_3\equiv Si-O^-$ and $O_3\equiv Si\cdot$ and capture of a hole to $O_3\equiv Si^+$ and $O_3\equiv Si\cdot$. The presence of the charged sites suggested by Hubner (187) are not consistent with the XPS data. Rather, the experimental evidence pointed to the existence of trivalent silicon, $O_3\equiv Si\cdot$ coordinated by water in its neutral state (hydrogen bonded) and also to non-bridging oxygen. In this last model both interfacial strain and oxygen deficiency are implicated in hole trap formation,

since the strained layer which has been shown to exist in the first 15-30 Å of the oxide is also a region of low oxygen content with a mixed composition of SiO, Si_2O and Si_2O_3 (143).

4A.3.6 Other models.

There are several models for positive charge and interface state generation during irradiation or electron injection (188-191). Analogous models could be considered as potential candidates to explain the similar degradation occurring under negative BTS. A key factor in most of these models is the release of some neutral species in the bulk as a result of energy loss to the oxide from either the radiation or hot electrons. This species then diffuses to the interface where it reacts with Si-H bonds, strained Si-O-Si bonds or other defects generating positive charge and interface states. Atomic hydrogen (188,191) and excitons (188,190) have been the main contenders for the diffusing species. Although there is no evidence to suggest that a significant electron current is flowing under BTS conditions to enable release of any species in the bulk, the similarity of the observed degradation under all types of stress makes these models worth noting. The possible role of hydrogen or water in causing degradation is also a recurring theme in the literature.

4B: RESULTS AND DISCUSSION

4B.1 Kinetics of positive charge generation.

Although the kinetics of the negative bias instability have already been widely studied, it is apparent from the previous discussion that the discrepancy in the various results makes these difficult to interpret. Many of the studies, also, were carried out ten to twenty years ago and given the sensitivity of this phenomenon to changes in processing, it seemed worthwhile to repeat some of these measurements using present day technology. The remainder of this section (4B.1) therefore contains observations concerning the phenomenology and kinetics of $Q_{ot}(+)$ generation. The relationship of this charge to hole trapping is covered in 4B.2 and interface state generation is considered in 4B.3. The high frequency and quasi-static C-V method was used to determine interface state distributions in depletion and the quasi-static and ideal curves were used for the rest of the band gap.

4B.1.1 Hysteresis in C-V curves and loss of $Q_{ot}(+)$.

Following negative BTS, the C-V curve is shifted to negative voltages and also shows some hysteresis due to the generation of slow-states (Fig 4.4). The magnitude of this hysteresis and indeed the precise position of the C-V curve changes with successive voltage sweeps. This is because a portion of the positive charge generated during negative BTS is easily lost. Application of a positive voltage (as occurs during a C-V sweep) or elevation of the temperature is sufficient to neutralise or de-trap this charge. To obtain an accurate measure of $Q_{ot}(+)$, it is important, therefore, to use the first C-V curve taken after BTS and to sweep from negative to positive voltages. Some charge may be removed even then, since Breed reported the annealing of positive charge generated by BTS, even at room temperature and zero bias (66). However, provided that C-V curves are always taken immediately after the stress, this effect should be both

small and consistent.

4B.1.2. Use of ΔV_{mg} to measure oxide charge.

Throughout the present work, the midgap voltage shift, ΔV_{mg}, has been used as a measure of the change in oxide charge alone, i.e. assuming that there was no contribution from interface states to the shift at this point. As mentioned previously, there will be no net charge in the interface states at midgap if all the states above midgap are acceptors and all those below are donors. The reverse condition, where all the states above midgap are donors and all those below are acceptors, can give the same result but requires the additional assumption that there is also an equal number of states above and below midgap. If either of these conditions holds, then V_{mg} is a neutral point for interface states and ΔV_{mg} represents the change in oxide charge alone. Several authors have assumed this to be the case for interface states generated by different types of stress: avalanche hole injection (47), negative BTS (48,49), high field stress (48) and irradiation (48,50,172). From C-V measurements, it is impossible to prove conclusively that V_{mg} is a neutral point for interface states or to distinguish whether the states are donors or acceptors over any particular part of the band gap. However, as will now be shown, the assumption that at V_{mg} there is no net charge in BTS-induced interface states was consistent with the data obtained in the course of this work.

Fig.4.4 shows the shift of the high frequency C-V curve following application of a negative BTS of -4MV/cm for 1 h at 250°C to a 28 nm oxide with an aluminium gate (curve B). On reversing the stress (+4 MV/cm for 1/2 h at 250°C) the curve moves back to more positive voltages (curve C). Comparison of avalanche hole injection measurements on samples subjected to (a) negative BTS, (b) negative BTS followed by positive BTS or (c) unstressed, showed that all the positive charge generated by negative BTS is removed by a subsequent positive BTS. These experiments are described in detail in sections

4B.2.2 and 4B.2.3. Following negative BTS and subsequent positive BTS, there is considerable interface state generation, however, causing distortion of the C-V curve. Nevertheless, curve C in Fig. 4.4 crosses the initial curve (curve A) very close to V_{mg}. Hence, the net charge in the interface states is aparently zero at midgap and V_{mg} can be used as a measure of the change in oxide charge alone.

4B.1.3 Location of $Q_{ot}(+)$

If the positive oxide charge, $Q_{ot}(+)$, resides close to the $Si-SiO_2$ interface, the ΔV_{mg} due to this charge should increase linearly with oxide thickness for a given BTS, since:

$$\Delta V_{mg} = Q_{ot}(+)x_o/\varepsilon_{ox}$$

where x_o is the charge centroid, measured from the metal-oxide interface. It can be seen from Fig.4.5 that this relationship holds for three sets of differently processed wafers with oxide thicknesses up to 50 nm.

One suggested cause of $Q_{ot}(+)$ is hole trapping (cf.4A.3.1). To see if the location of this charge is compatible with the filling of oxide hole traps, photo I-V measurements were carried out before and after avalanche hole injection into an MOS capacitor. This showed intrinsic hole traps to be located at both interfaces. Typical photo I-V curves measured before and after avalanche injection of 6.25 x 10^{15} holes into cap. B on wafer A11 are shown in Fig.4.6. Since there is distortion on both sides of the I-V curve following hole injection, the trapped holes can be estimated to be within 5 nm of each interface. Any charge trapped near to the metal gate would have no effect on the C-V curve while charge trapped in hole traps near to the silicon would cause a ΔV_{mg} which increased linearly with oxide thickness. It is possible therefore, that the generation of positive oxide charge during negative BTS is due to the filling of intrinsic hole traps.

4B.1.4 Time dependence of ΔV_{mg}.

Similar variation of ΔV_{mg} with time was observed to that found by other workers. The results could be fitted to either a ln t or a t^n relationship. Wafer MN6, for example, showed a good fit to $\Delta V_{mg} \propto$ ln t (Fig.4.7), whereas the wafers of run U could better be fitted to $\Delta V_{mg} \propto t^n$, where n depends on the stress temperature and varied from 0.28 to 0.33 over the temperature range studied (Fig.4.8). These latter results are similar to those reported by Shiono et al. (63) for their oxides grown in HCl, although no HCl was used in the inner tube during oxidation of run U. A ΔV_{mg} versus ln t plot for these samples (using the same data as that in Fig.4.8) is shown in Fig.4.9. It can be seen that although in the latter the data do not fit quite so well to a straight line, the scatter is within experimental error. The functions $\Delta V_{mg} \propto$ ln t and $\Delta V_{mg} \propto t^n$ are very similar and in many cases the time dependence cannot easily be distinguished. In one case only, that of boron doped polysilicon gate capacitors, was there clearly a dependence of ΔV_{mg} on ln t whereas a plot of ln $\Delta V_{mg} \propto$ ln t deviated from linearity at longer times (Fig.4.10).

4B.1.5 Temperature dependence of ΔV_{mg}.

The build-up of positive charge under negative bias is obviously strongly influenced by temperature as is the subsequent annealing of this charge under positive or zero bias. A good fit was obtained to an Arrhenius law, $\Delta V_{mg} \propto \exp(-E_a/kT)$ with an activation energy, E_a of between 0.24 and 0.43 eV. The value varied from batch to batch but was fairly reproducible (±0.03 eV) for wafers processed simultaneously. Neither the substrate dopant (n- or p-type) nor the dopant level had a significant effect. (See Table 4.1). An Arrhenius plot for one of the wafers from Table 4.1 (MN6) is shown in Fig.4.11. The activation energies found in this work are lower than the values reported in the early literature on the negative bias instability (0.6-1.5 eV - see section 4A.1.2.2.). However, Haller et al. recently reported a very low value of 0.18 eV (48). The latter may

well be artificially low due to the sample not being cooled under bias. Nevertheless, this parameter would appear to vary quite considerably between samples.

Since water or related species are implicated in several models for the negative bias instability, it is interesting to compare the above activation energies for $Q_{ot}(+)$ generation with those for the diffusion of H_2O, -OH and other species in SiO_2. From experiments using tritiated steam as the oxidising ambient, it appears that water incorporated during oxide growth does not redistribute under BTS (192,193). It therefore has no charge associated with it and/or it is not mobile. However, water subsequently diffused into the oxide will be in a chemically different environment. It has been suggested that under certain conditions it may act as a positively charged species (194-196), possibly bound H^+. Protons from ethanol have been shown to be exchanged with SiO_2 giving some support to this idea (197).

The activation energy for the diffusion of water into SiO_2 films is around 0.3 eV (196) which is in the range of values found in this work for the activation energy of $Q_{ot}(+)$ generation (and also the same as that found in ref. (49) for interface state generation.) It is known to diffuse as two separate entities but the nature of these has not been established. E_a for the diffusion of water in bulk SiO_2 (fused silica) is somewhat higher at about 0.6 eV (196). Of similar magnitude are the activation energies for steam oxidation, oxidation in wet O_2 and for the out-diffusion of H_2O from steam grown oxides (196). 0.6 eV is closer to some of the activation energies for $Q_{ot}(+)$ generation previously measured (66,67,167). E_a has also been measured for the diffusion of molecular hydrogen and of -OH in SiO_2 (198). These values were 0.49 eV and 0.68 eV respectively. H_2 has been suggested as a possible product of Si-H bond breaking during BTS and 0.49 eV is again fairly close to the activation energies for positive charge generation measured in this work.

If the diffusion of some water-related species is involved in positive charge generation during negative BTS, several factors could be responsible for the widely varying magnitude of the associated E_a. These include differences in the structure of the SiO_2 itself e.g.the existence and size of micropores and the water content of the lattice, differences in the nature of diffusing species and in the way in which it binds to the lattice. However, the diffusion process would be expected always to involve the making and breaking of hydrogen bonds so that E_a should remain of the order of the strength of such a bond i.e. between 0.2 and 1 eV (199).

4B.1.6. Field dependence of ΔV_{mg}.

For all the wafers examined an $E^{3/2}$ dependence of ΔV_{mg} on the field was found. This is illustrated in Fig.4.12 for wafer MN6. A similar relationship was recently reported by Haller et al. (48). As yet however, there is no explanation in terms of mechanism or kinetics for this striking form of the field dependence.

4B.1.7 Electrode dependence of ΔV_{mg}.

There is a strong dependence of ΔV_{mg} on the gate electrode. Aluminium gate capacitors were generally used in this work since polysilicon gate devices showed much smaller shifts for a given time and magnitude of stress. Times of the order of 10^3 h are usually required to measure significant degradation of polysilicon capacitors (63) whereas similar degradation could be achieved with aluminium gate devices in a matter of minutes. The thickness of the aluminium gate electrode and the means of its deposition were also found to be important. This is discussed further in the following sections.

One batch of boron-doped polysilicon gate devices showed somewhat larger ΔV_{mg} shifts than expected. These were notably larger than similarly processed phosphorus-doped polysilicon devices (Fig.4.9). This difference was seen whether the boron was introduced by thermal

doping or ion-implantion. For the ion-implanted wafers, ΔV_{mg} was found to decrease with increasing temperature of post-deposition anneal. This is in agreement with the work of Sinha et al. (168) who found an increase in positive BTS shift and a decrease in negative BTS shift following high temperature H_2 anneal. They attributed this to enhanced boron penetration of the oxide and subsequent negative charging of some boron-related species under stress.

4B.2 Positive charge generation and hole trapping.

4B.2.1 Correlation between ΔV_{mg}(BTS) and ΔV_{mg}(AI).

While conducting measurements on capacitors with metal gate electrodes of different thicknesses, it was observed that samples showing a smaller ΔV_{mg} during negative BTS (ΔV_{mg}(BTS)) also showed a smaller shift during avalanche hole injection (ΔV_{mg}(AI)). This is illustrated in Table 4.2. ΔV_{mg}(AI) after injection of 2.5 x 10^{15} holes/cm^2 is compared to ΔV_{mg}(BTS) following a stress of 3 MV/cm for 1h at 250°C for three wafers with differing aluminium electrode thicknesses. The shifts are normalised to an oxide thickness of 28 nm assuming all the charge to be resident at the Si-SiO_2 interface. For both negative BTS and hole injection, the trend is the same. A thinner gate electrode resulted in a smaller ΔV_{mg}. This suggests that hole trapping might be responsible for the negative bias instability and further experiments were performed to confirm this. These experiments are discussed below while the dependence of hole trapping on metal gate thickness is considered further in 4B.2.6.

4B.2.2 ΔV_{mg}(BTS) and filling of intrinsic hole traps.

Avalanche hole injection was carried out on a sample which had undergone negative BTS and the variation of ΔV_{mg} and V_{mg} was compared with that measured during hole injection on a virgin sample. If ΔV_{mg}(BTS) were entirely due to hole trapping, a subsequent hole

injection experiment would show the same saturation value of the absolute midgap voltage, V_{mg}, as is observed on a virgin sample. On the other hand, ΔV_{mg}, the midgap voltage shift, occurring during hole injection would differ in the two cases by the magnitude of ΔV_{mg} already found during negative BTS. This was, in fact, observed experimentally. Fig.4.13a shows ΔV_{mg} during hole injection on a virgin sample and on one having undergone a stress of 3 MV/cm at 250°C for 5.5 h. The difference between the saturation values of the two curves tends towards the value of ΔV_{mg} already present after BTS and at the moment hole injection commenced; in this case about -2 V. Fig.4.13b shows the absolute value of the midgap voltage in the two cases and both curves saturate at the same level.

4B.2.3 Removal of positive charge by positive BTS.

It has already been shown in 4B.1.2 and Fig.4.4 that positive BTS after negative BTS returns V_{mg} to a value very close to its initial value. Subsequent avalanche hole injection resulted in a ΔV_{mg} versus injected charge curve almost identical to that obtained on an unstressed sample (Fig.4.14). Hence, the positive stress removed all the holes from the oxide; no new hole traps were created during BTS at either polarity and there was no change in the existing capture cross sections as a result of filling and emptying of the traps. It would seem, therfore, that hole trapping during BTS is a completely reversible process. It does, however, result in considerable interface damage which will be discussed in 4.3.

4B.2.4 Capture cross sections of hole traps.

All the avalanche hole injection curves which were measured could be fitted using first-order kinetics to two exponentials i.e. to give two capture cross sections and effective trap densities. Table 4.3 shows the capture cross sections and effective trap densities measured for three samples. Two of these were unstressed, having thin (15 nm) and thick (1 μ) aluminium gates, respectively. The third was taken

from the same wafer as the unstressed device with a 1 μ gate but was first subjected to a stress of -3 MV/cm for 1/2 h at 250°C. Comparing samples 2 and 3 in Table 4.3, it can be seen that BTS resulted in the filling of the larger capture cross section trap, with $\sigma = 3 \times 10^{-14}$ cm^2. This trap was not seen in the sample with 15 nm aluminium as the gate. Instead a trap with $\sigma = 6 \times 10^{-16}$ cm^2 was found at relatively low effective density (8×10^{11} $/cm^2$). Wafers with thin and thick gates both showed a trap with $\sigma = 5 \times 10^{-15}$ cm^2, but this had a lower density in the thin metal gate samples.

4B.2.5 Current flow during negative BTS.

No measurable current flow was observed during negative BTS once a steady temperature had been reached. During heating a small peak was sometimes seen at low temperature (<100°C) presumably due to Na^+ drift (inset Fig.4.15). Above 200°C the current again began to rise somewhat but as soon as the temperature stopped increasing, the current dropped rapidly to zero (Fig.4.15). It is difficult to decide whether this current is due to ionic effects such as K^+ drift or whether it is a displacement current associated with hole trap filling. The current increase does occur at about the temperature expected for K^+ ion motion (200); also, integration of the current over the time of the applied stress cannot account for the observed ΔV_{mg}. Heating to 305°C at 2 MV/cm and immediately cooling again to room temperature resulted in a shift of about -1 V in the C-V curve for a 28 nm oxide of capacitor area 1.1×10^{-2} cm^2. This is approximately equivalent to 10^{-5} C/cm^2 of injected charge, if it is all located at the $Si-SiO_2$ interface. However, only about 5×10^{-7} C/cm^2 of injected charge was found from integration of the current during the time of the applied stress (Fig.4.15). Filling of hole traps very near to the interface would not result in any displacement current which possibly could account for this discrepancy.

4B.2.6 Effect of gate electrode on hole trapping.

To investigate the dependence of hole trap densities and capture cross sections on the thickness of the metal gate, avalanche hole injection was conducted on a batch of similarly processed wafers with aluminium electrodes ranging from 15 to 1000 nm thick. ΔV_{mg} as a function of injected charge, Q_{inj}, is shown for these wafers (S1-S6) in Fig.4.16 after hole injection at 1.5×10^{-8} A/cm^2. After 8.5×10^{14} holes had been injected, there was 6 V difference between the shifts measured on S1 (1000 nm Al) and on S6 (15 nm Al). Figure 4.17 is a plot of ΔV_{mg} versus aluminium thickness (d_{Al}) and Fig.4.18 a plot of ΔV_{mg} versus $d_{Al}^{(0.25)}$. The latter gives a good straight line fit to the data but the reason for this is unknown.

Table 4.4 shows the capture cross sections (σ_i) and effective densities (N_{effi}) of hole traps obtained for each sample by fitting the hole injection curves. The total effective density of traps for each wafer is also shown. In all cases three traps gave the best fit to the data. The trap with the largest capture cross section σ_1 was least affected by the electrode thickness, showing no change in the value of σ. There is a tendency for trap density to increase with increasing gate electrode thickness for traps σ_1 and σ_2 and there is also an increase in σ for traps σ_2 and σ_3. If the effective trap densities are summed there are almost twice as many traps in the oxide with a 1000 nm Al gate as in that with a 15 nm Al gate.

The reason for the increasing density of intrinsic hole traps with increasing aluminium gate thickness may be that the magnetron sputtering technique introduces some radiation damage. This would be greater for longer sputtering times and hence thicker metal gates. However, the fact that thin aluminium gate capacitors show even fewer hole traps than polysilicon gate devices which were not exposed to radiation means that other factors are also at work. (Hole injection curves for polysilicon, thin and thick aluminium gate devices are

shown in Fig.4.19 with ΔV_{mg} normalised to a 28 nm oxide.) The most likely reason for the high hole trap density with polysilicon gates is that there is always a high temperature anneal following polysilicon deposition and this is known to increase hole trap density (186).

The effects of the electrode material and metal deposition technique were further examined using flash evaporation as well as magnetron sputtering to deposit aluminium, either directly on to the oxide or on top of a previously deposited polysilicon layer. Polysilicon electrodes without an aluminium contact layer were also used. Table 4.5 shows ΔV_{mg} after 10^{15} holes had been injected into similar wafers, differing only in their gate electrode. The wafer with thick flash-deposited aluminium (R10) electrodes showed far fewer hole traps than the one with magnetron sputtered aluminium of a similar thickness (S1). Wafer R12, with 425 nm of flash-deposited Al gave the lowest ΔV_{mg} followed by R10 (850 nm flash) and S6 (15 nm magnetron sputtered aluminium). Polysilicon gates resulted in almost twice the shift shown by the latter devices (10 V). 1 μ of magnetron sputtered aluminium gave a still higher shift and worst of all was polysilicon with flash-deposited aluminium on top of it. These results suggest that some radiation damage is probably occurring in the magnetron sputtering system, but polysilicon deposition is almost as detrimental from the point of view of hole trap creation. However, while the polysilicon gate devices show the presence of many hole traps in an avalanche injection experiment, these are less easily populated in a BTS experiment than the traps in the aluminium gate devices. This may be due to their being located further from the interface or simply at a different energy level.

Another factor, which may be important in determining the observed density of hole traps, is the effect of the gate electrode on the interfacial stress. Zekeriya and Ma (201,202) saw a decrease in both ΔV_{mg} and ΔD_{it} during irradiation for increasing thickness of the aluminium electrode. They were able to explain this using the bond strain gradient model (181-183), since thicker aluminium layers

relieve more effectively the intrinsic tensile stress at the $Si-SiO_2$ interface (Fig.4.20). When avalanche hole injection was done on devices with flash-evaporated aluminium of different thicknesses, a somehat similar trend of ΔV_{mg} with metal thickness was found. These results are illustrated in Fig.4.21(a) while those of ref. (201) are shown in Fig.4.21(b) for comparison. In Fig.4.21(b), ΔV_{mg} at first drops quickly with increasing aluminium thickness and then saturates. In Fig.4.21(a) ΔV_{mg} also decreases sharply but there is slight increase again at greater metal thicknesses. Since the overall trend is similar, however, it seems likely that interfacial stress is indeed a significant factor in causing hole trap generation, along with oxygen deficiency and exposure to radiation.

4B.2.7 Discussion of models for hole trapping.

Although it has now clearly been established that the positive charge generated during negative BTS is due to hole trapping, two basic questions remain to be answered. Firstly, what is the precise mechanism of hole trap filling and secondly, what is the physical nature of a hole trap? Either of the trapping mechanisms suggested in the literature, i.e. thermally activated tunneling of electrons from donor states in the oxide to the silicon conduction band (65) or thermal emission of holes in the silicon valence band directly into oxide hole traps (66) could be operating. Likewise, few of the models offering a physical description of the negative bias instability can be ruled out, since bond breaking (58) may well be the net result of hole trapping, while oxygen vacancies (68), VAP defects (187), Si-H, Si-OH (186) or strained Si-O-Si bonds (181-183) are all possible hole trapping sites. Nevertheless, some indications are available both from the literature and from the present work as to how the hole trapping process might occur. These will now be discussed.

Firstly, it seems clear that hole traps are related to non-stoichiometry of the oxide and in particular to oxygen deficiency. There is, for example, a strong correlation between the number of hole

traps and exposure to a high temperature anneal in N_2. Conversely, a short final POA in O_2 substantially reduces hole trapping (186). This agrees with the findings of the present work where polysilicon gate devices, which always receive a high temperature POA in N_2, have a large number of hole traps. $O_3{\equiv}Si\cdot$ $\cdot Si{\equiv}O_3$ and $O_3{\equiv}Si\text{-}O\cdot$ $\cdot Si{\equiv}O_3$ generated by removal of oxygen or of H_2O from the lattice have been suggested as likely hole trapping sites (186). Using XPS, Grunthaner et al. (143) found experimental support for the existence of trivalent silicon, $O_3{\equiv}Si\cdot$, and negatively charged non-bridging oxygen at the interface after exposure to irradiation by high energy electrons, a process known also to generate hole traps. These authors propose that bond breaking is occurring at strained Si-O-Si bonds near to the interface resulting in generation of the $O_3{\equiv}Si\cdot$ $\cdot O\text{-}Si{\equiv}O_3$ site. This site may subsequently capture either an electron or a hole which is compatible with the evidence that hole traps are in fact amphoteric sites able also to act as shallow electron traps (186). The XPS data mentioned above were not consistent with the presence of the dipolar site suggested by Hübner (187) in the VAP model.

It is interesting to consider the value of the activation energy for hole trapping measured in this work (around 0.3 eV) which is close to that expected for breaking of a hydrogen bond and certainly compatible with the trapping of holes at $O_3{\equiv}Si\cdot$ $\cdot Si{\equiv}O_3$ or $O_3{\equiv}Si\cdot$ $\cdot O\text{-}Si{\equiv}O_3$ sites coordinated by water. In the XPS spectrum measured in ref. (143), there was a shift of 0.3 eV in the oxygen 2p signal of the neutral non-bonding oxygen species suggesting that it is indeed hydrogen bonded.

The role of stress in the interfacial region is still unclear. It has been demonstrated that a strained layer exists in the first 15-30 Å of the oxide, the width and precise nature being determined by processing conditions (181-183). This layer is also essentially oxygen deficient with regions corresponding to SiO and Si_2O. It seems likely that whatever their nature, the density, energy and position of hole traps will be affected by the interfacial stress.

One possible reason why hole traps in polysilicon gate capacitors are not readily filled by BTS, although avalanche hole injection shows them to exist with a high density, is that the different interfacial stress causes them to be located further from the interface or at a higher energy than those in aluminium gate devices. There is also evidence from the dependence of hole trapping on the thickness of flash-evaporated aluminium electrodes that stress is playing a role.

If, as seems likely, hole traps are associated with singly occupied non-bonding orbitals on either oxygen or silicon these may be generated in several different ways. Higher energy holes entering the oxide by avalanche hole injection or being generated as a result of irradiation are probably able to break weak Si-O-Si bonds as has been suggested (143). However, it may be that the reason for the lower positive charge generation in a BTS experiment is that only already broken or the most highly strained bonds may be accessed by this low energy technique. These bonds must also be suitably located in position as well as energy. If a sample is irradiated, even if the positive charge is annealed out, it will have a higher density of hole traps under avalanche injection or BTS conditions than an unirradiated sample. Likewise, a sample having undergone hole injection and then having this charge neutralised is more sensitive to positive charge generation by subsequent BTS. No traps are generated or populated in the BTS experiment though that would not also be reached by the higher energy avalanche injection technique.

The fact that no current flow was observed during negative BTS suggests that either trap hopping or thermally activated tunneling is indeed the most likely mechanism for injection of holes into the insulator. The process of positive charge generation may, therefore, be quite different from that occurring during current injection under avalanche or F-N tunneling conditions. Unfortunately, the kinetics of hole trapping during BTS do not reveal as much as might be expected about the mechanism by which hole trapping occurs. The time and field dependence of the positive charge generation is mainly useful in

predicting degradation under less severe device operation conditions. The $E^{3/2}$ dependence of ΔV_{mg} on the field is striking but does not immediately suggest a model. The time dependence is not so clearly established but $\Delta V_{mg} \propto \ln t$ seems to give the best fit to all the data. This relationship could be indicative of either a direct hole hopping or an electron emission process, however.

4B.3 Interface state generation accompanying hole trapping and detrapping or neutralisation of trapped holes.

It is already well known that hole trapping results in interface degradation (47,148). Interface state generation accompanying hole trapping during both negative BTS and avalanche hole injection has been examined and in this section the results will be presented and compared with those in the literature. Neutralization or detrapping of holes by positive BTS after negative BTS was also found to influence the interface state density distribution. The resulting degradation was similar to that seen following irradiation or F-N injection, both of which supply both holes and electrons to the oxide.

4B.3.1 $\Delta D_{it}(mg)$ during negative BTS.

Interface state generation across the whole bandgap was observed following negative BTS. As has been reported by other workers (49), the change in interface state density at midgap, $\Delta D_{it}(mg)$, was found to be proportional to ΔV_{mg} i.e. to the number of trapped holes. This is illustrated in Fig.4.22(a). From this plot, the ratio of interface states generated to trapped holes was found to be 1.36 ($N_h = \Delta V_{mg}\varepsilon_{ox}/d_{ox}q$ and $N_{it} = D_{it}(mg) \times 1.1$ eV). This is somewhat larger than the value of one reported in reference (49). This one-to-one ratio has also been reported for interface state generation by ionising radiation (178). Although a long period was required before all the interface states were created; the limiting value only being reached after a year.

As previously mentioned, some of the positive charge created by negative BTS is very easily detrapped, by the taking of a C-V curve, for example and the apparently high ratio measured here may be partly due to an underestimation of ΔV_{mg}. However, in ref. (49) the samples were cooled with the gate floating which would be expected to lead to an even greater loss of charge (and hence to a real N_{it}/N_h ratio of less than one.) It may also be that some samples are simply more sensitive to interface state generation as a result of hole trapping than others.

4B.3.2 $\Delta D_{it}(mg)$ during avalanche hole injection.

Lai reports a linear relationship between holes trapped and $\Delta D_{it}(mg)$ during avalanche hole injection (47) (from his data N_{it}/N_h-0.35). This linear relationship was confirmed by the present measurements. It can be seen from Fig.4.22(b) that $\Delta D_{it}(mg)$ saturates with injected charge but is a linear function of ΔV_{mg}, i.e. of the number of trapped holes. A higher ratio of interface states to trapped holes was found in the present work than was reported by Lai, however. N_{it}/N_h is close to 2 from the data of Fig.4.22(b). This ratio is also greater than that found for negative BTS. During avalanche injection a substantial hole current flows through the oxide. Traps may well be filled further from the interface than is the case for negative BTS where no current flow was seen under steady-state conditions. Holes trapped at, say, 50 A into the oxide would cause substantially less shift in a 280 A oxide than those trapped at the interface. This may be one of the reasons for the higher N_{it}/N_h ratio found for hole injection compared with that for negative BTS. Experiments on slightly thicker oxides would help to decide this question. Alternatively, more of the holes entering the oxide and filling hole traps may be de-trapped again under the somewhat higher field conditions existing in the case of avalanche injection. The interface degradation would remain or even be increased by de-trapping of the holes, although the measured ΔV_{mg} would be reduced, thereby increasing the apparent N_{it}/N_h. That Lai

measured a much smaller N_{it}/N_h ratio for avalanche hole injection than was found here suggests that there is substantial sample-to-sample variation in this value. In neither ref. (47) nor this work was any evidence found for a 1:1 ratio of interface states and trapped holes though.

4B.3.3 D_{it} distribution after negative BTS/positive BTS.

Positive BTS after a negative BTS resulted in removal of the positive oxide charge and generation of a peak in the interface state distribution at between 0.15 eV and 0.25 eV above the middle of the silicon band gap (Fig.4.23). The precise position varied somewhat from batch to batch and to a lesser extent (±0.02 eV) with the parameters chosen for the C-V analysis. The magnitude of the peak was dependent on the severity of the initial negative BTS if the subsequent positive BTS was kept constant. Figure 4.24 shows, qualitatively, the effect on the peak size of increasing the stress time during negative BTS. The peak area was found to be linearly proportional to the amount of positive charge removed by the positive BTS (ΔV_{mg}^{+}). This is illustrated in Fig.4.25. The peak could not be generated by a positive BTS alone; once generated, however, it could be reversibly removed and regenerated by negative BTS and positive BTS, respectively.

Using avalanche hole injection followed by photo-injection of electrons, Lai has shown that such a peak is created in the interface state density distribution by a two-stage process involving the capture of electrons by trapped holes (47). This sequence of measurements was also performed in the course of the present work and these results were confirmed (Fig.4.26). It appears that a similar two-stage process of interface state generation occurs during hole trapping and neutralisation or detrapping in the negative BTS/positive BTS sequence. However, if Fig.4.23 is examined carefully it can be seen that as well as the peak being generated following positive BTS there is a corresponding reduction in D_{it} beyond 0.25 eV above

midgap. It is possible, therefore, that the peak is due to a redistribution of states in energy rather than to the actual generation of new states. Maintaining the sample at 250°C under zero bias for about 45 min. (after negative BTS then positive BTS) resulted in flattening out of the peak. Again, this may be a further redistribution effect rather than an actual anneal of states.

4B.3.4 D_{it} distribution generated by F-N tunneling or irradiation.

Other processes which supply both electrons and holes to the oxide, such as high field stress and irradiation, have also been reported to generate the 0.2 eV peak (48). Confirmation that electron injection by F-N tunneling generates this peak is given by the results presented in Fig.4.27. The peak height is apparently proportional to the injected charge (Fig.4.28).

Haller et al. (48) conducted a study comparing the rate of D_{it} generation with ΔV_{mg} following negative BTS, F-N tunneling and irradiation. They observed a peak at 0.1 eV below midgap following negative BTS alone and at 0.15 eV above midgap after irradiation or high field stress. They report precisely the same linear relationship between the heights of these peaks and ΔV_{mg} for all three forms of stress. Their data are summarised in Fig.4.29. The measurements in this work did not support the existence of such a relationship. Firstly, following negative BTS, no peak was observed at 0.1 eV below midgap but there was a fairly uniform increase across the whole band gap. More importantly, there were difficulties with the use by these authors of the so-called corrected midgap voltage shift, '$\overline{\Delta V_{mg}}$', for the electron injection experiments. Electron trapping in the oxide partly obscures the positive charge build-up which also accompanies high field stressing. In ref. (48) it was assumed that this could be corrected for by taking the slope of the initial negative ΔV_{mg}, close to time zero, as being due solely to positive oxide charge. This assumption is not valid since there is no reason to suppose that

electron trapping does not begin immmediately and that the early ΔV_{mg} is not due to the net effect of both electron trapping and positive charge generation with the latter predominating. This is demonstrated in Fig.4.30, where ΔV_{mg} versus injected charge is plotted for different values of the current density. It can be readily seen that a different initial slope is obtained for each current density. The measurement interval chosen may also affect the equilibrium between positive and negative charge in the oxide. The use of $\overline{\Delta V_{mg}}$ as a measure of the positive trapped oxide charge during electron injection does not seem justifiable, therefore.

That the height of the peak in the interface state density, ΔD_{it}(peak), plotted versus $\overline{\Delta V_{mg}}$ is a straight line is not surprising since it has already been shown that ΔD_{it}(peak) is proportional to the injected charge, Q_{inj} (Fig.4.25) and Haller et al. themselves confirm this relationship. If $\overline{\Delta V_{mg}}$ is deliberately made proportional to Q_{inj} then ΔD_{it}(peak) must also vary linearly with $\overline{\Delta V_{mg}}$. Although there are undoubtedly qualitative similarities in the degradation occurring during irradiation, BTS and F-N tunneling, the evidence for the existence of the same quantitative relationship between ΔV_{mg} and interface state generation for these three techniques (as expressed in Fig.4.26) does not seem to be very firmly based.

4B.3.5 D_{it} distributions measured by charge pumping.

Since calculation of the interface state distribution from C-V curves requires measurement of the surface potential, it can be affected by any lateral non-uniformities of the charge at the interface. It is possible that these might cause artifacts in the D_{it} distribution such as the peak observed above midgap. To check that this peak was genuinely due to interface states and not to non-uniformities, transistors and capacitors from the same wafer were similarly stressed. Charge pumping measurements were then conducted on the transistors and C-V measurements conducted on the capacitors.

The corresponding interface state densities were then calculated and compared.

Polysilicon gate devices were used in this case and so as to increase the rate of degradation a short high field stress was first applied (9.1 $MVcm^{-1}$ for 120 s). This was then followed by negative BTS (- $2MVcm^{-1}$ at 250°C for 1/2 h) and positive BTS (+2 $MVcm^{-1}$ at 250°C for 15 min.). It would be expected that, if anything, the high field stress would increase the charge non-uniformity at the interface.

The D_{it} distributions resulting from charge pumping and from C-V measurements are shown in Fig.4.31. It can be seen that the peak at 0.2 eV above midgap is present in both distributions. The charge pumping measurement however, resulted in a much sharper and narrower peak than the C-V measurement. A possible reason for this is that the charge pumping measurement is inherently more accurate in the regions near to the band edges where the peak occurs, whereas the C-V technique has greatest accuracy near to the centre of the band gap. On the other hand, comparison of D_{it} distributions from charge pumping and from C-V measurements with no structure in the distribution showed a good correlation between the two sets of data. Figure 4.32 shows such distributions for a capacitor and a transistor from the same wafer (CP1) (a) before stress and (b) after negative BTS at 2 MV/cm and 250°C for 1 h. This difference in shape between the interface state density peak calculated from C-V measurements and from charge pumping measurements needs further investigation.

4B.3.6 Nature of interface states generated by hole trapping.

Since the number of interface states generated during hole trapping is proportional to the number of holes trapped, it seems likely that these states are physically similar. It may be that the same defect generates either a hole trap or an interface state depending on its distance from the interface. From the previous

discussion on the nature of hole traps (4B.2.7), the amphoteric trivalent silicon and non-bonding oxygen site is the most likely candidate for this defect. It is already established (45) that $\equiv Si\cdot$ is responsible for the U-shaped distribution of interface states present in as-grown oxides. Interface state generation across the whole band gap during negative BTS and avalanche hole injection is compatible, therefore, with an increase in $\equiv Si\cdot$ being responsible for these states. $\equiv Si^{+}$ and $\equiv Si\cdot$ are probably present, although holes trapped near to the interface were shown to be easily de-trapped even at zero applied voltage, suggesting a tendency for the neutral state to predominate.

Positive BTS after negative BTS neutralises or detraps holes trapped in the oxide and creates a peak in the interface state density at 0.2 eV above midgap. The reversibility of the generation and removal of this peak on positive and negative BTS respectively, suggests that it is caused by a redistribution in energy of existing traps rather than the generation of new states. The cause of this redistribution may be either a re-orientation of bonds in space or a change in the charge state of the defect.

Wafer	E_a (eV)	Doping level
MN2 (n-type)	0.43	10^{15}
MN6 "	0.37	10^{17}
MP6 (p-type)	0.43	"
ST2 (n-type)	0.24	"

Table 4.1 Activation energies for the generation of $Q_{ot}(+)$ during negative BTS for n- and p-type wafers.

Al thickness (nm)	ΔV_{mg}(BTS) (3MV/cm:1h:250°C)	ΔV_{mg}(AI) (2.5 x 10^{15} holes/cm)
15	-0.2 V	-2.57 V
60	-0.3 V	-4.54 V
1000	-1.1 V	-8.04 V

Table 4.2 ΔV_{mg} following negative BTS and hole injection on capacitors of varying metal gate electrode thickness.

NEGATIVE BIAS INSTABILITY: RESULTS

Sample	σ_1, (cm^2)	N_1, ($/cm^2$)	σ_2, (cm^2)	N_2, ($/cm^2$)	σ_3, (cm^2)	N_3 ($/cm^2$)
15 nm Al (ST1)	-	-	5E-15	1E12	6E-16	8E11
1 µ Al (ST2)	3E-14	5E12	4E-15	3E12	-	-
1 µ Al (ST2) after BTS at -3 MV/cm 250°C 0.5 h	3E-14	3E12	6E-15	3E12	-	-

Table 4.3 Capture cross sections (σ_i) and effective hole trap densities (N_i) for wafers ST1 and ST2 before stressing and for ST2 after negative BTS.

	S1 (1000 nm)	S3 (250 nm)	S4 (125 nm)	S5 (50 nm)	S6 (15 nm)
$N_{eff}1$ ($\times 10^{12}$)	2.3	1.5	1.5	1.3	1.0
σ_1 ($\times 10^{-14}$)	1.5	3.3	3.0	2.3	2.3
$N_{eff}2$ ($\times 10^{12}$)	1.5	1.0	0.9	0.9	0.8
σ_2 ($\times 10^{-14}$)	5.2	0.1	0.1	0.6	0.6
$N_{eff}3$ ($\times 10^{12}$)	1.0	1.4	1.2	0.8	1.0
σ_3 ($\times 10^{-15}$)	3.8	8.4	6.2	0.8	0.6
TOTAL N_{eff}: ($\times 10^{12}$)	4.7	3.9	3.6	2.9	2.8

Table 4.4 Capture cross sections and effective densities of hole traps in wafers with different thicknesses of the aluminium gate electrode (S1,S3,S4,S5,S6).

Wafer	Electrode material	ΔV_{mg} after 10^{15} holes injected
R4	Poly 450 nm + 850 nm flash	-13.0 V
R5	Poly 450 nm + 425 nm flash Al	-11.0 V
R6	Poly 450 nm + 100 nm flash Al	-16.0 V
R7	Poly 450 nm	-10.0 V
R10	Flash Al 850 nm	-5.3 V
R12	Flash Al 425 nm	-4.0 V
R13	Flash Al 100 nm	-8.0 V
S1	Sputt. 1 μ Al	-12.0 V
S6	Sputt. 15 nm Al	-5.5 V

Table 4.5 ΔV_{mg} after injection of 10^{15} holes into samples with different gate electrodes. Run R has 58 nm oxide and run S has 51 nm oxide. Shifts have been normalised to 58 nm of oxide, assuming all charge to be at the interface.

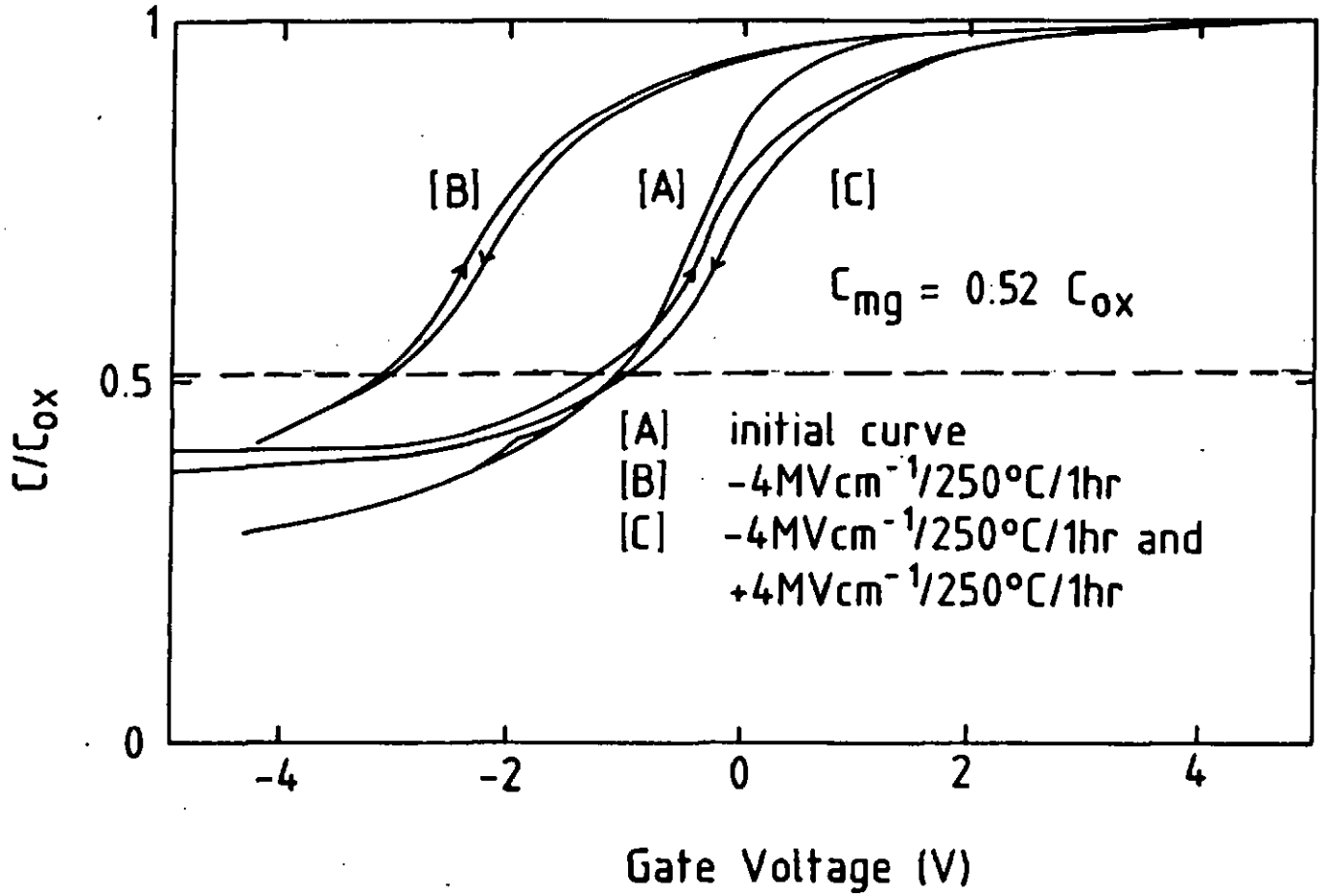


Figure 4.4 High frequency (1MHz) C-V curves for wafer ST2, cap F (A) before stress (B) following negative BTS and (C) following negative plus positive BTS.

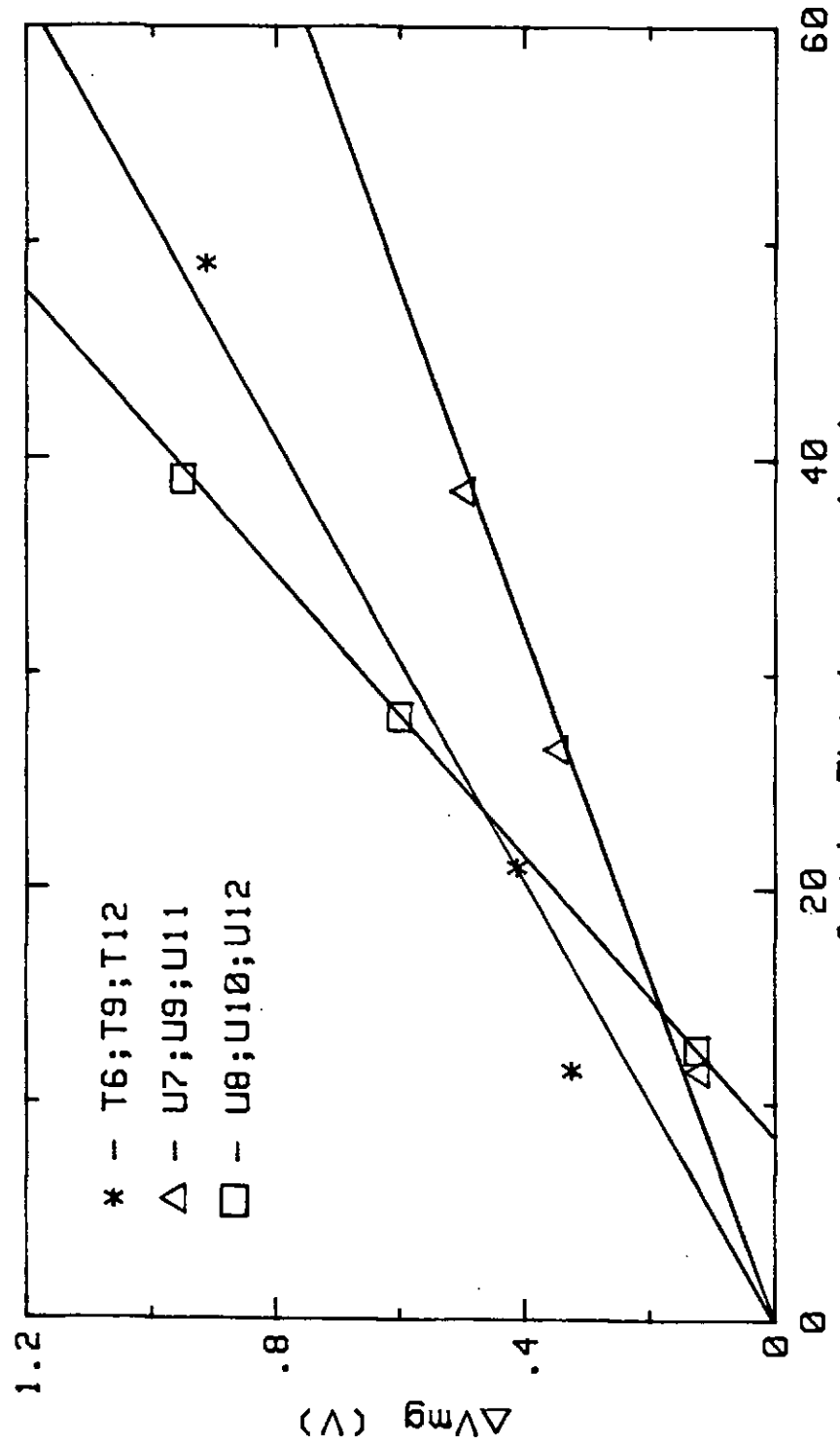


Figure 4.5 Relationship between ΔV_{mg} and oxide thickness for three sets of differently processed wafers.

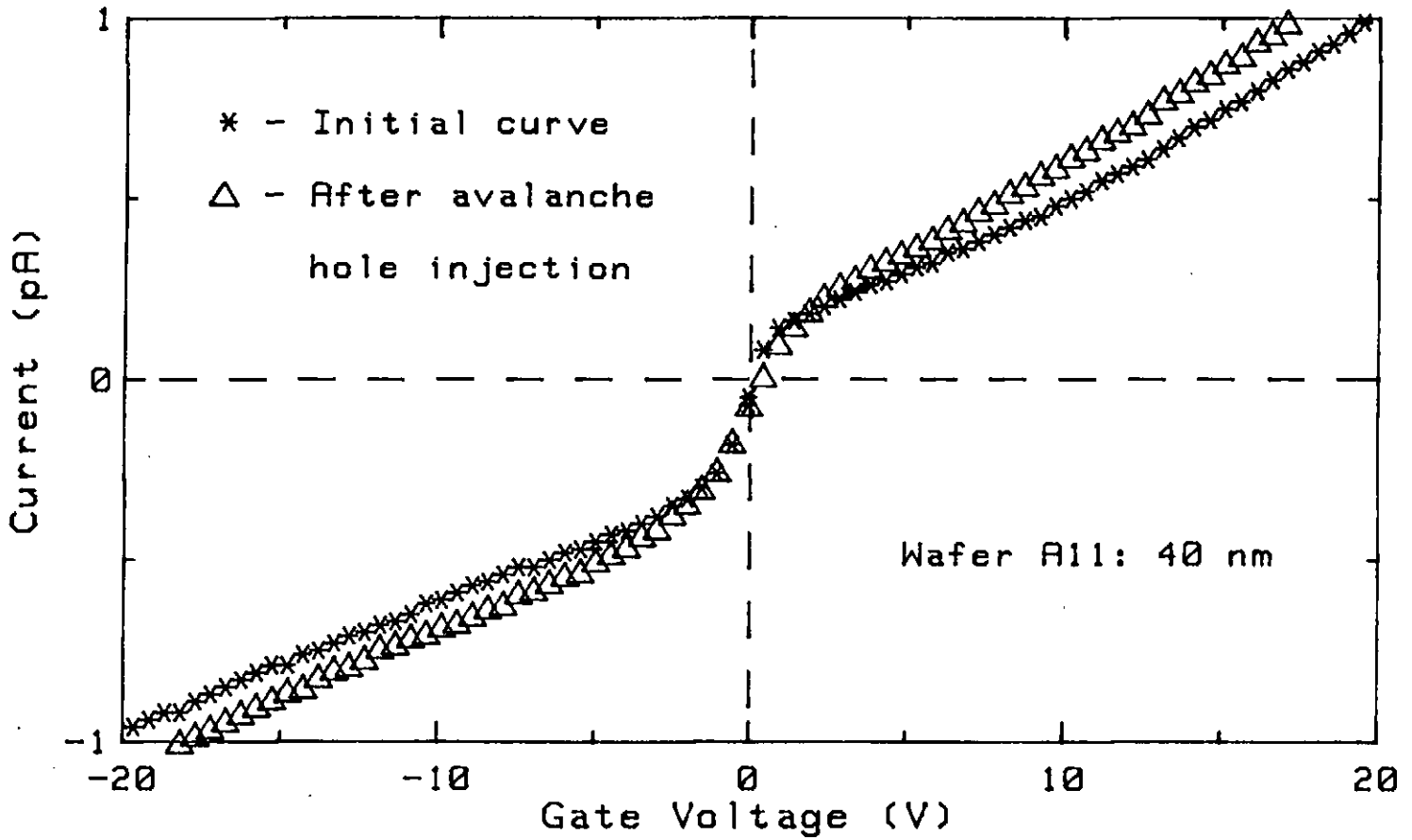


Figure 4.6 Photo I-V curves taken before and after photoinjection of 6.25×10^{15} holes cm^{-2} on wafer A11, cap F.

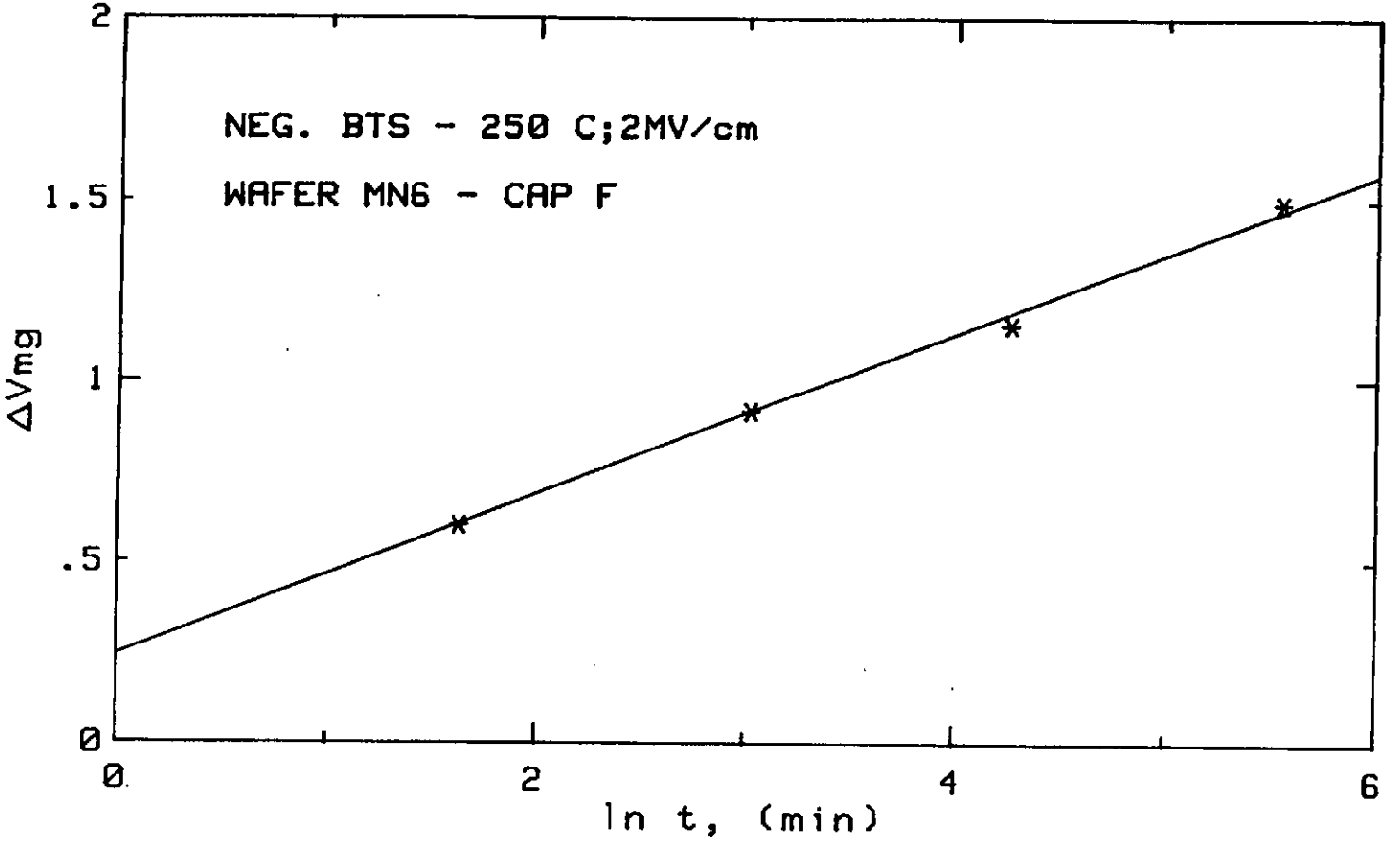


Figure 4.7 - ΔV_{mg} versus ln t during negative BTS at 250°C and 2MV/cm on wafer MN6, cap.F.

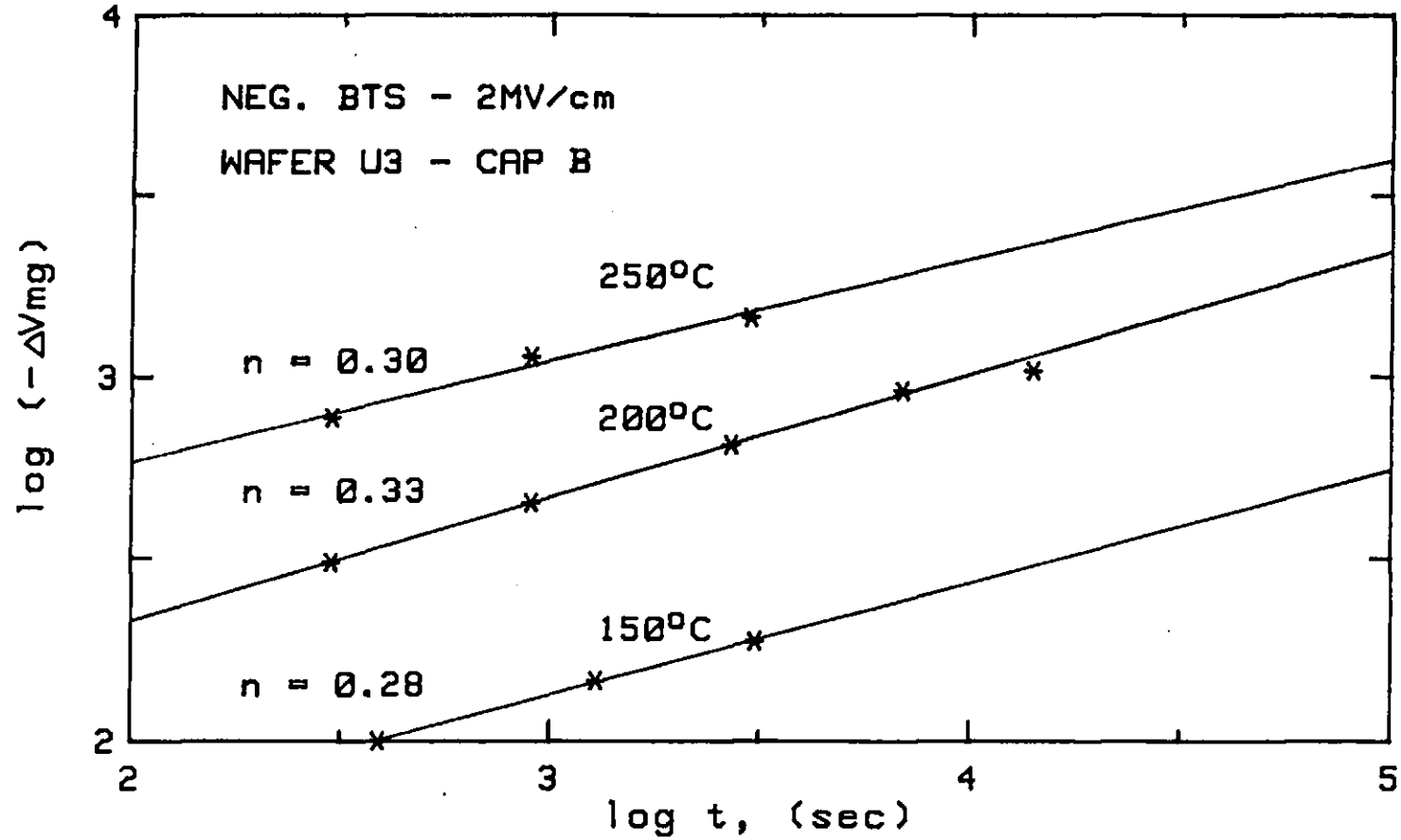


Figure 4.8 Log($-\Delta V_{mg}$) versus log t for wafer U3, cap.B during negative BTS at 2MV/cm and various temperatures.

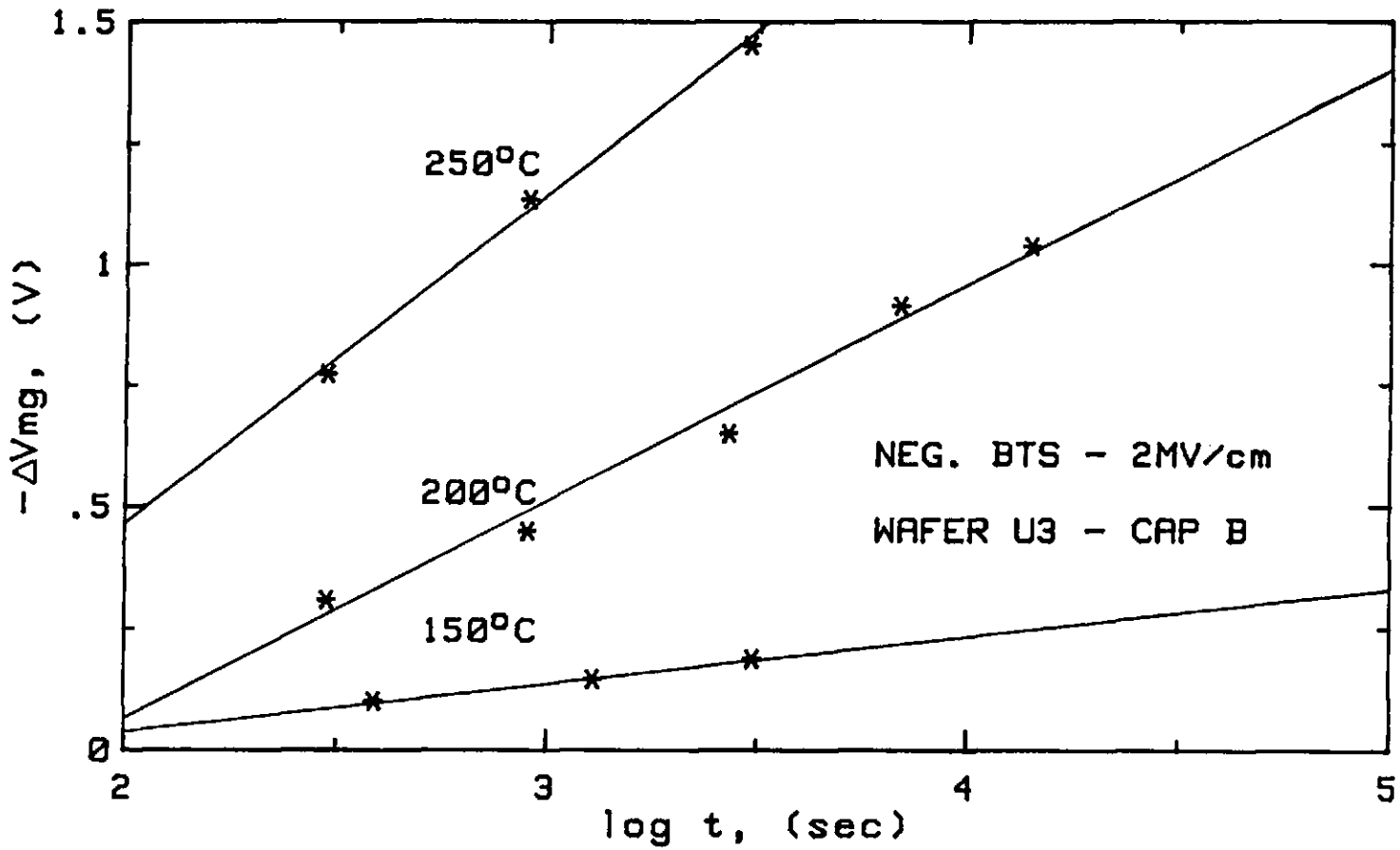


Figure 4.9 $-\Delta V_{mg}$ versus log t for wafer U3, cap.B during negative BTS at 2MV/cm and various temperatures.

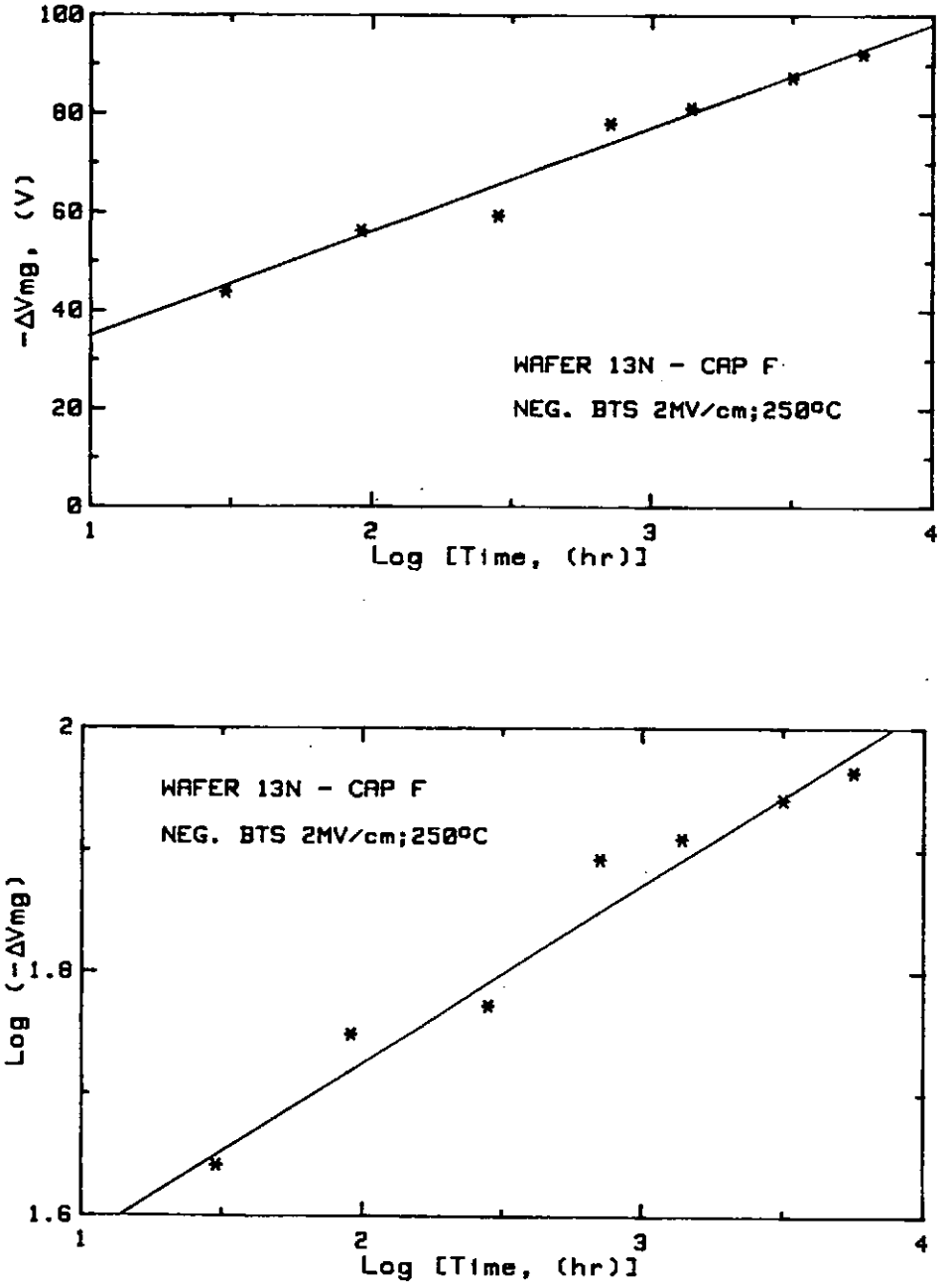


Figure 4.10 (a) $-\Delta V_{mg}$ versus log t

(b) Log $(-\Delta V_{mg})$ versus log t for wafer 13N, cap F during negative BTS.

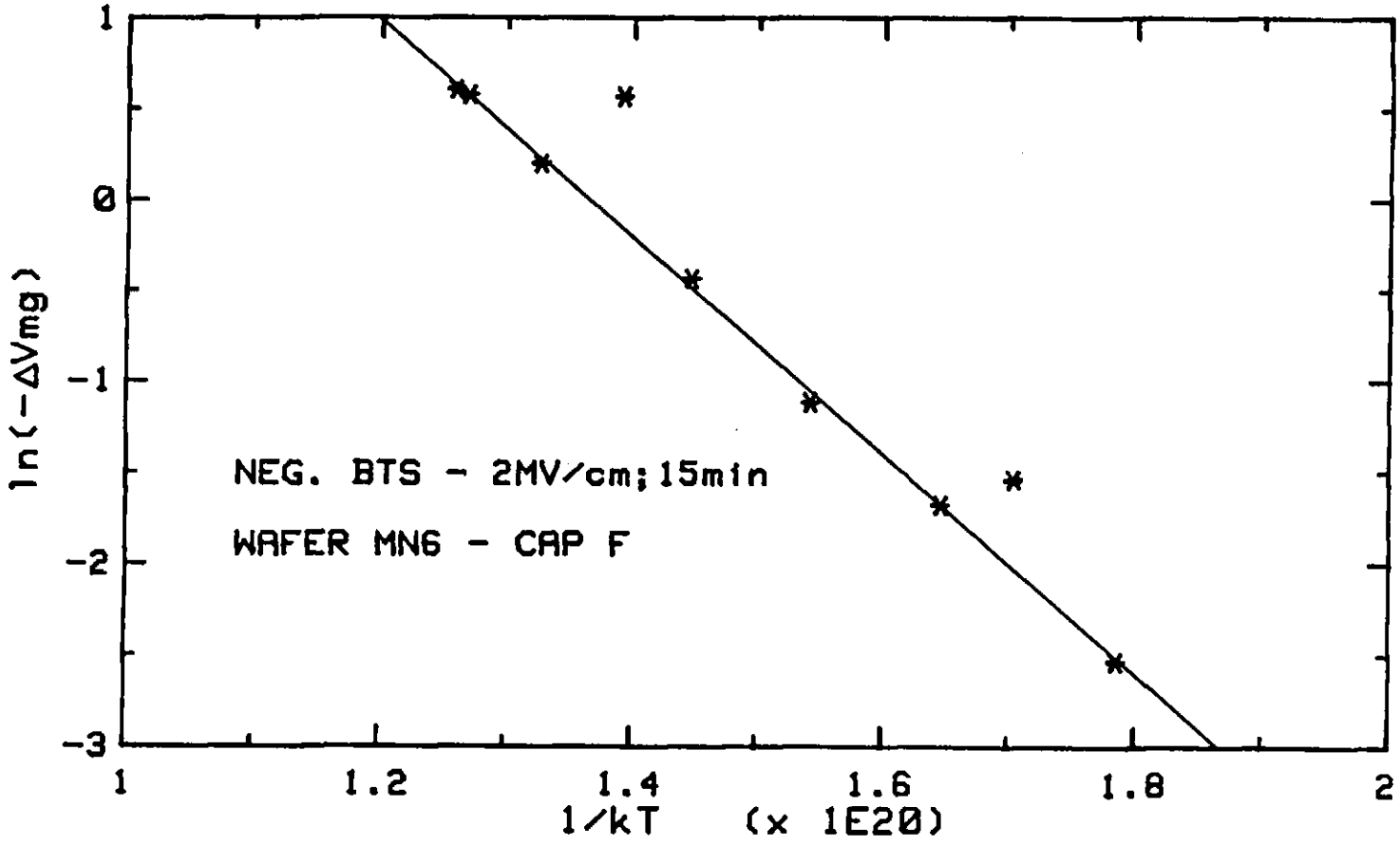


Figure 4.11 Arrhenius plot for the variation in ΔV_{mg} with temperature during negative BTS on wafer MN6, cap.F at 2MV/cm for 15 min.

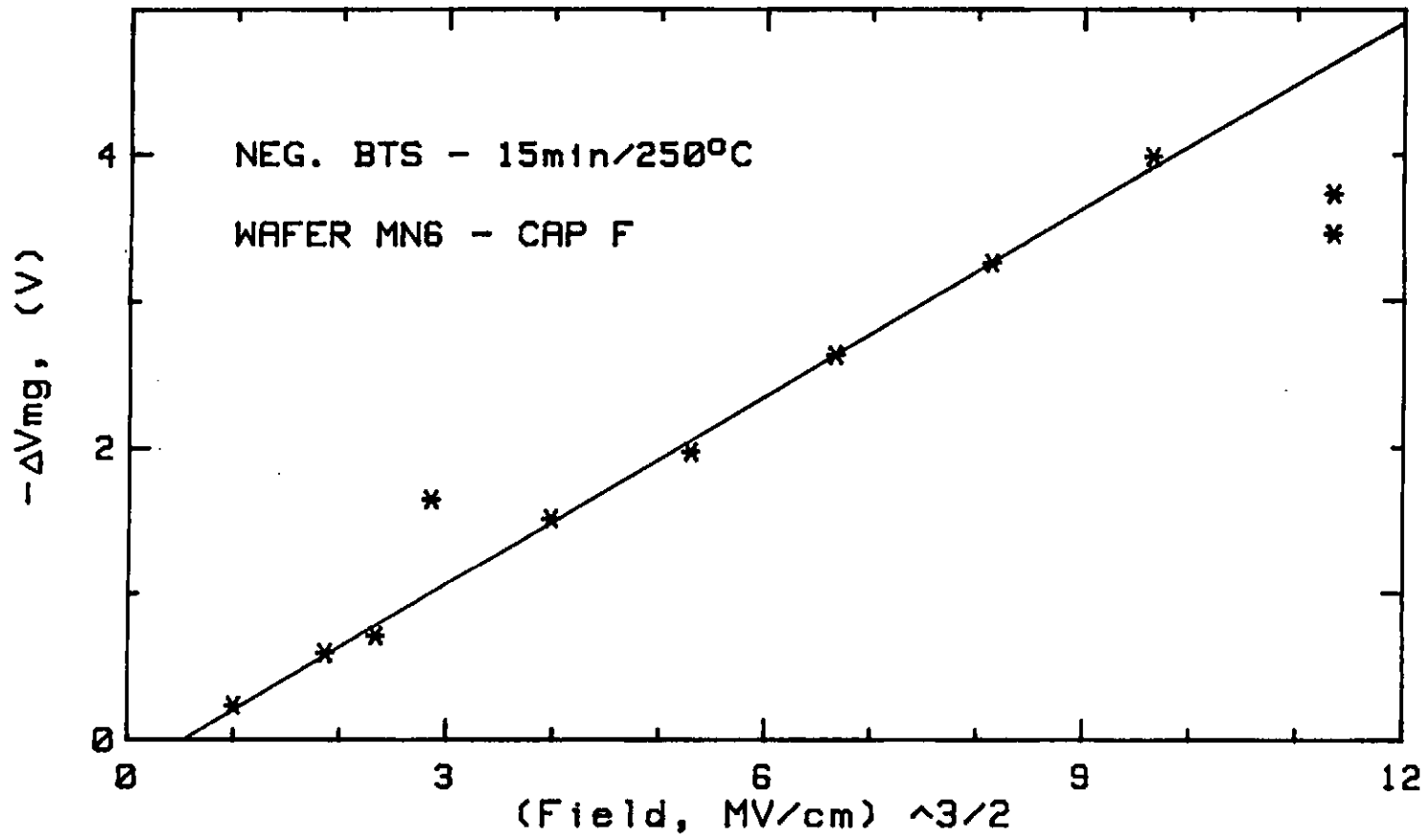


Figure 4.12 - ΔV_{mg} versus $E_{ox}^{3/2}$ for negative BTS on wafer MN6, cap.F at 250°C for 15 mins.

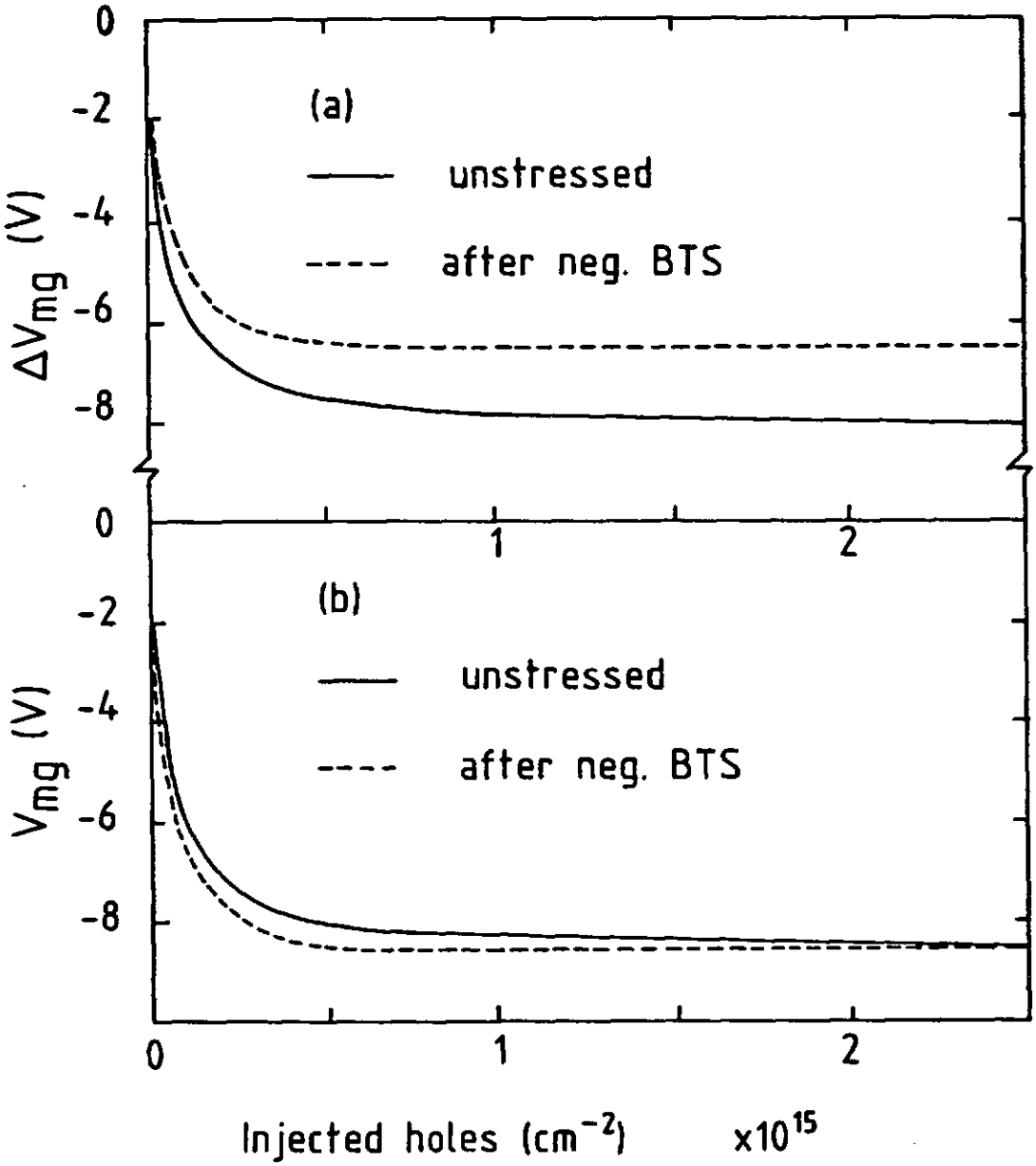


Figure 4.13 (a) ΔV_{mg} and (b) V_{mg} as a function of injected holes on an unstressed sample and after negative BTS at 3MV/cm and 250°C for 5½ hours. (wafer ST2-cap F).

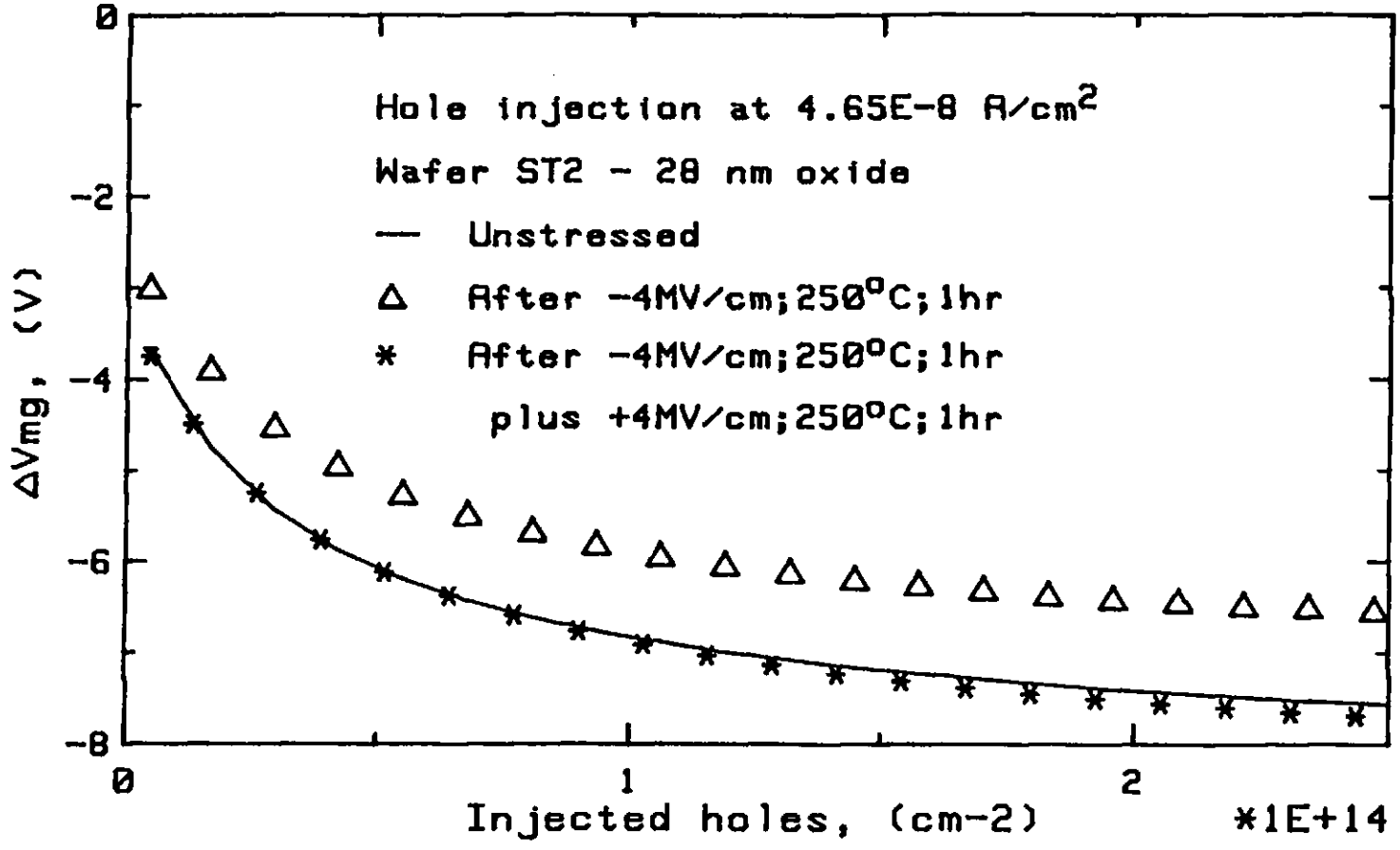


Figure 4.14 ΔV_{mg} during avalanche hole injection on an unstressed sample, after negative BTS and after negative plus positive BTS on cap.F wafer ST2.

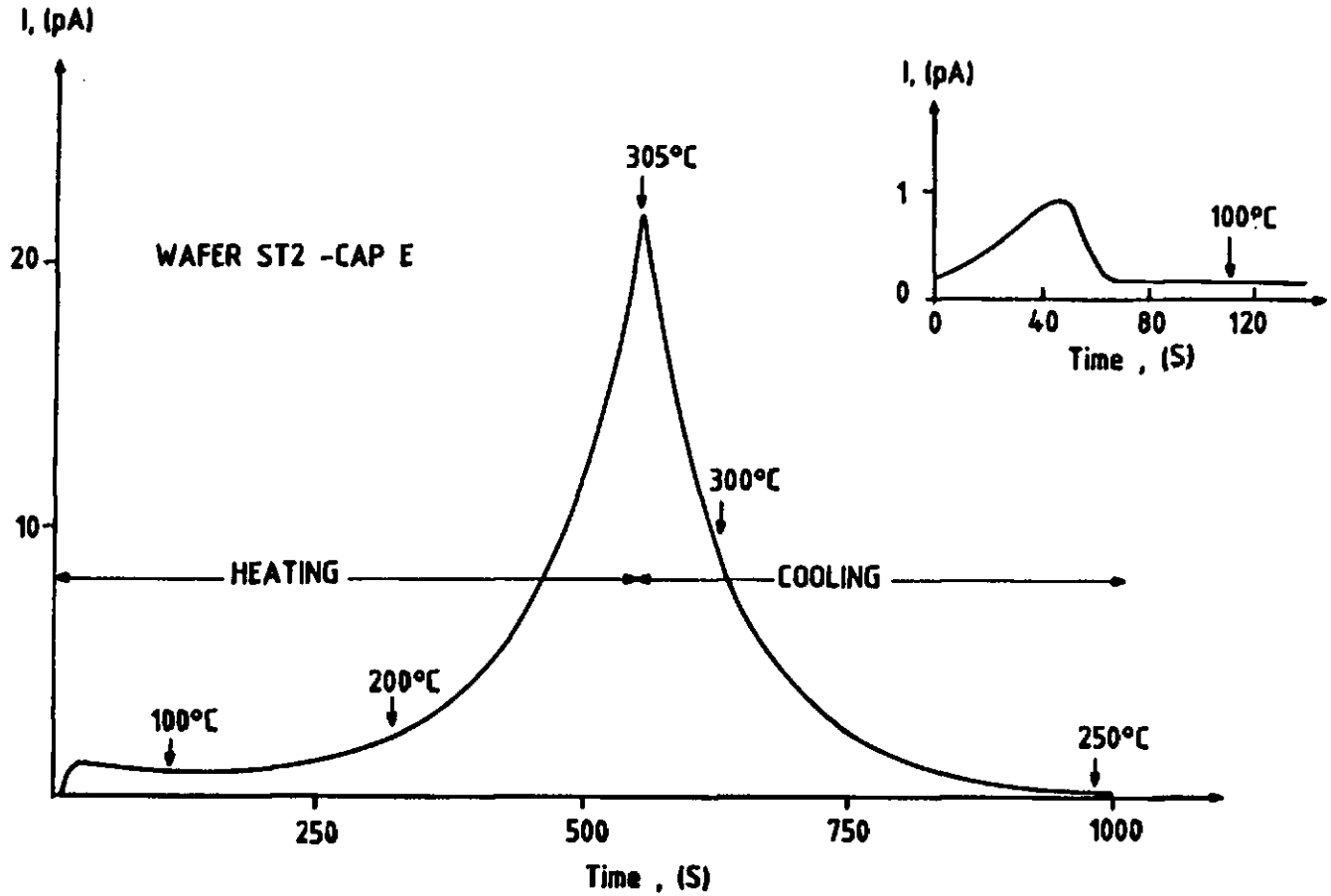


Figure 4.15 Current flow under negative bias (2MV/cm) during heating and cooling of cap.E on wafer ST2.

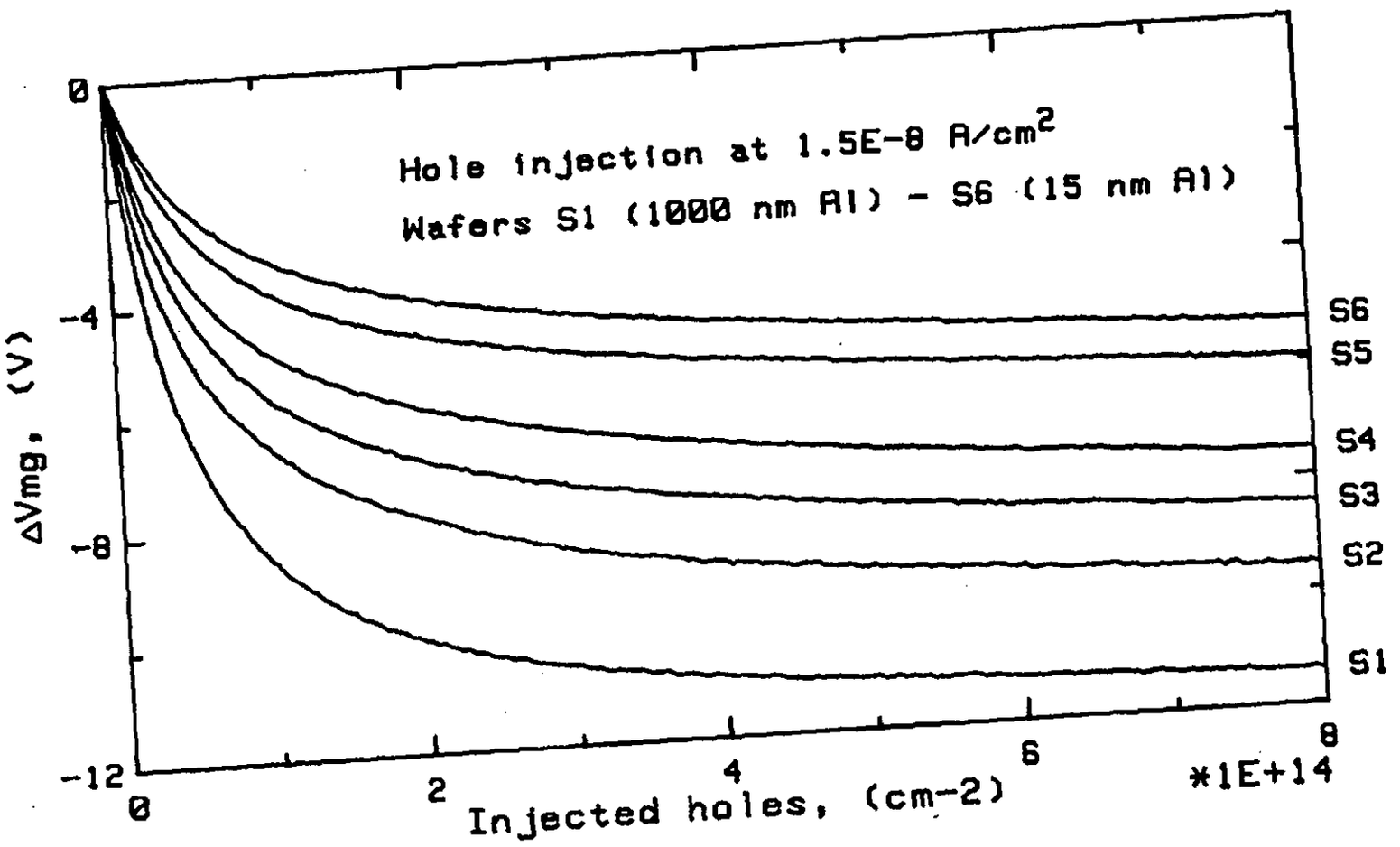


Figure 4.16 ΔV_{mg} during avalanche hole injection on wafers with different thicknesses of the aluminium gate electrode.

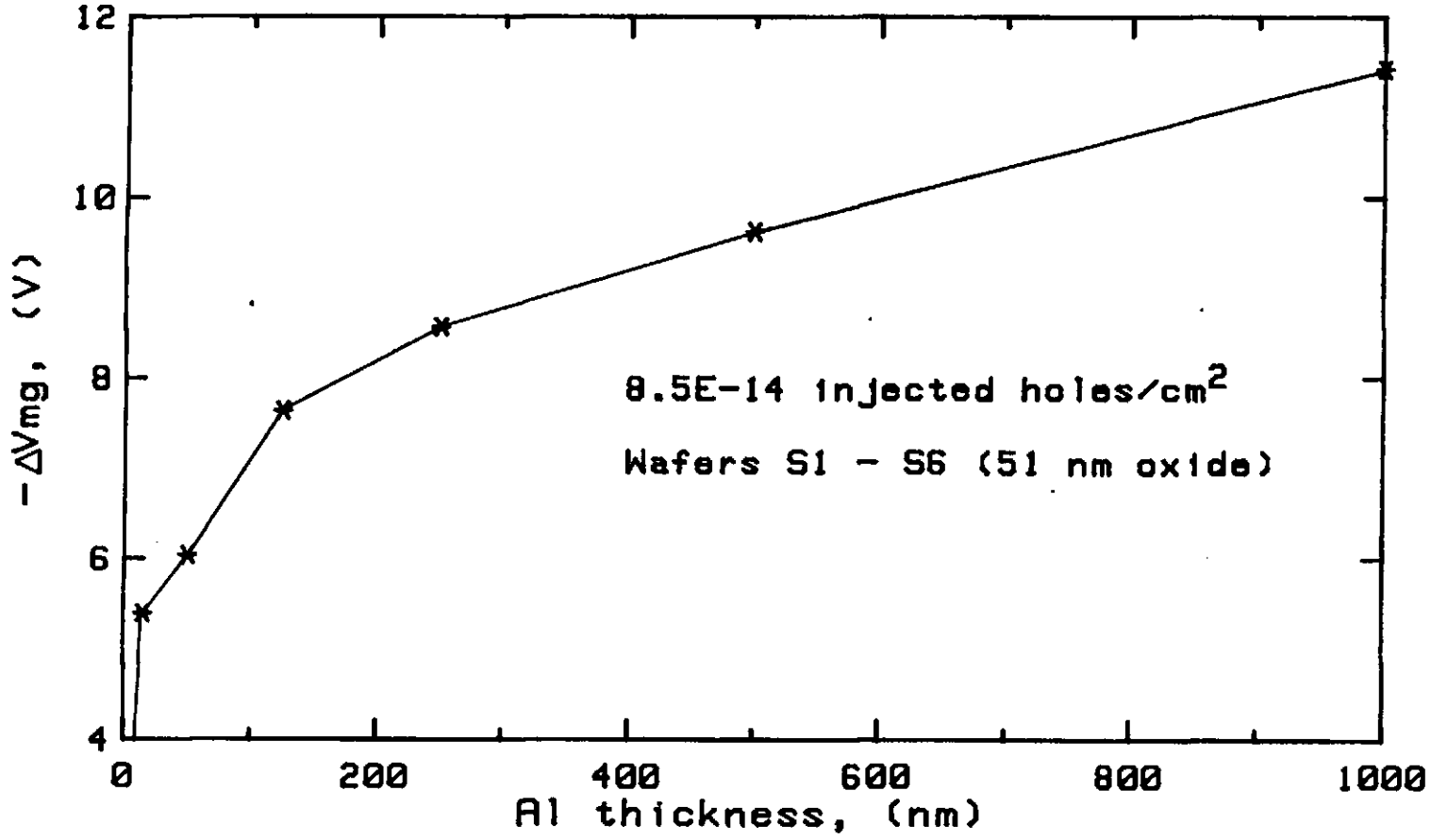


Figure 4.17 - ΔV_{mg} after injection of 8.5×10^{-14} holes cm^{-2} as a function of the thickness of the aluminium gate electrode.

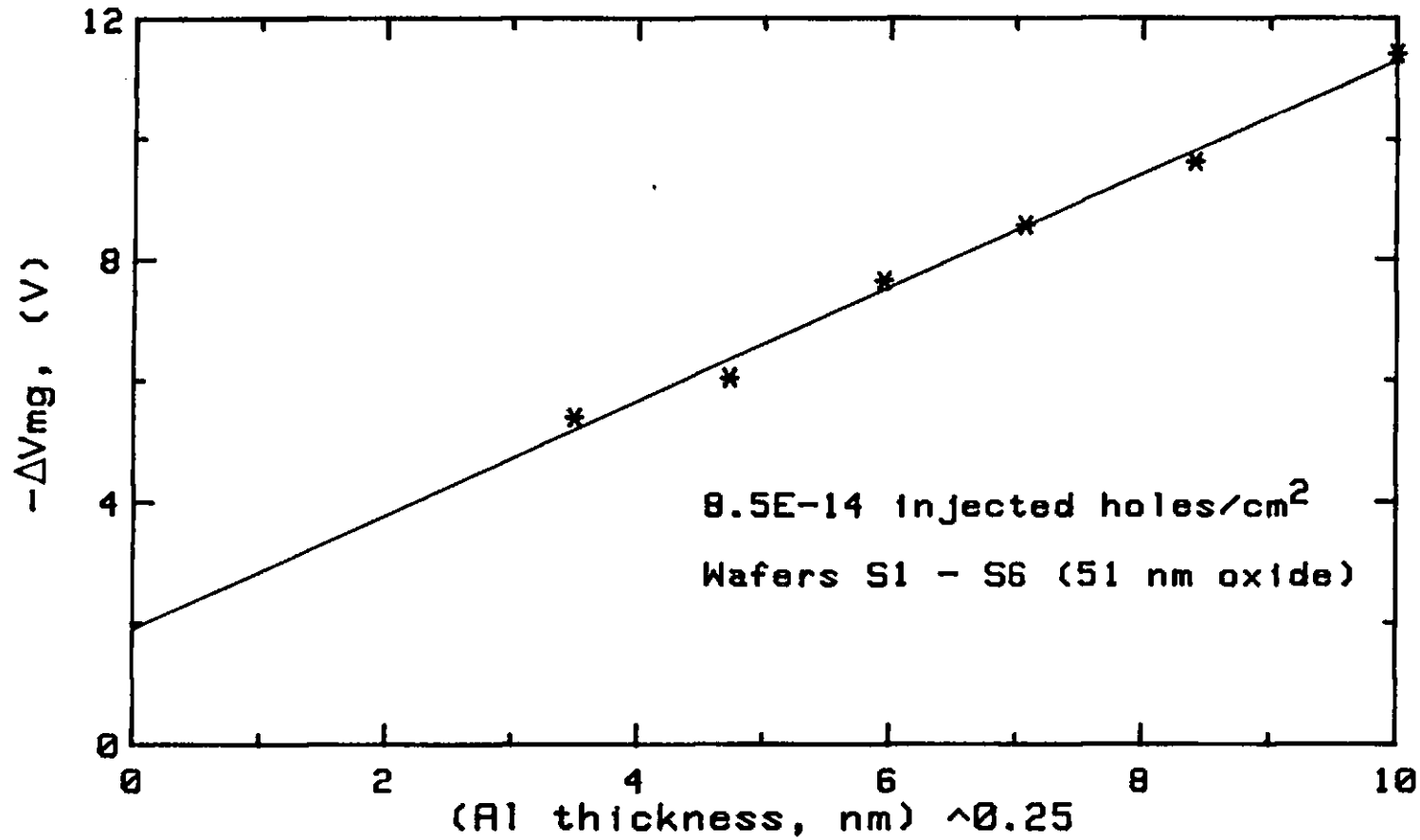


Figure 4.18 $-\Delta V_{mg}$ after injection of 8.5×10^{-14} holes cm^{-2} as a function of the fourth root of the aluminium gate electrode thickness

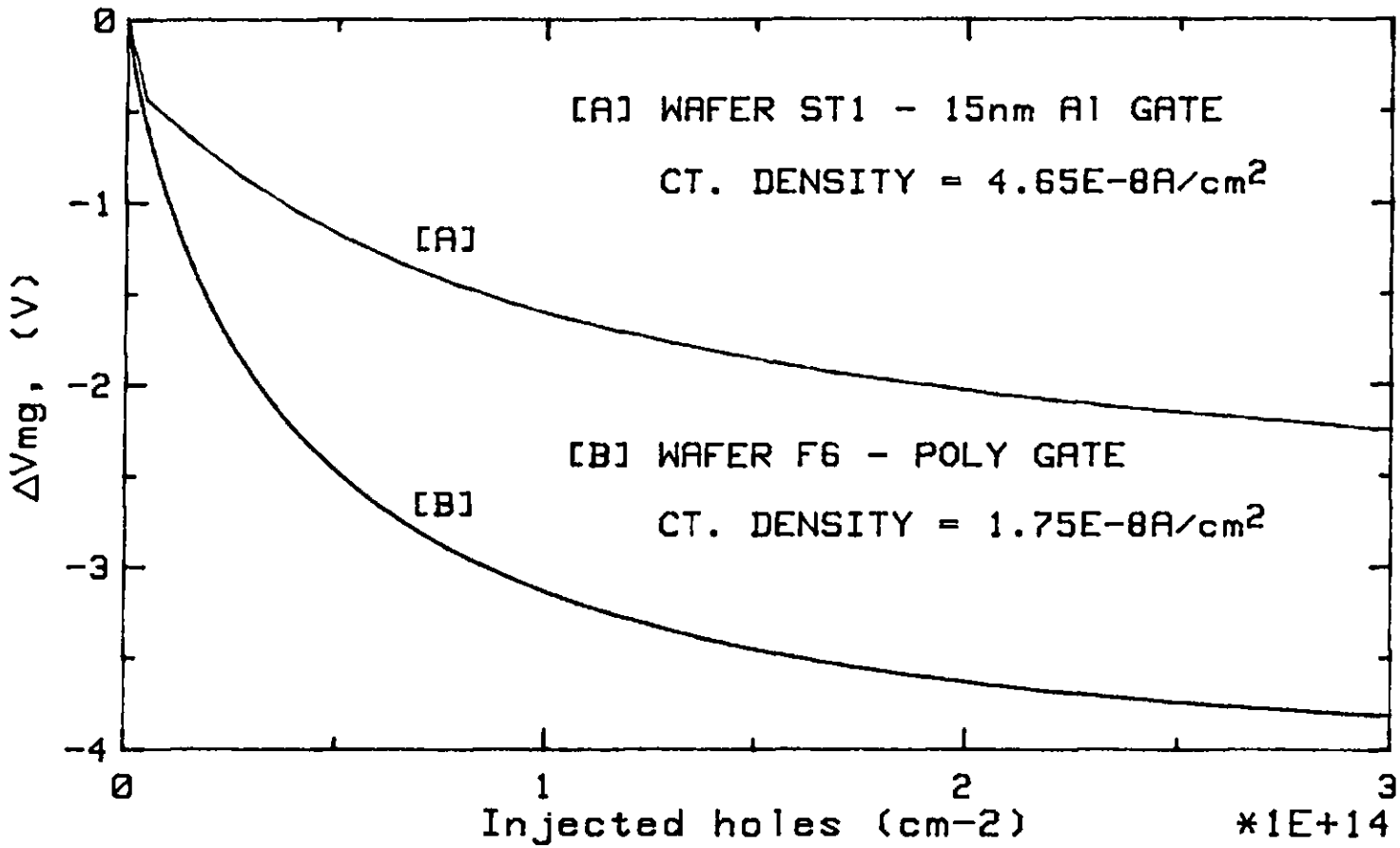


Figure 4.19 ΔV_{mg} during avalanche hole injection on polysilicon gate (F6) and thin aluminium (15nm) gate (ST1) capacitors. (The shifts have been normalised to an oxide thickness of 28nm assuming all charge to be located at the interface)

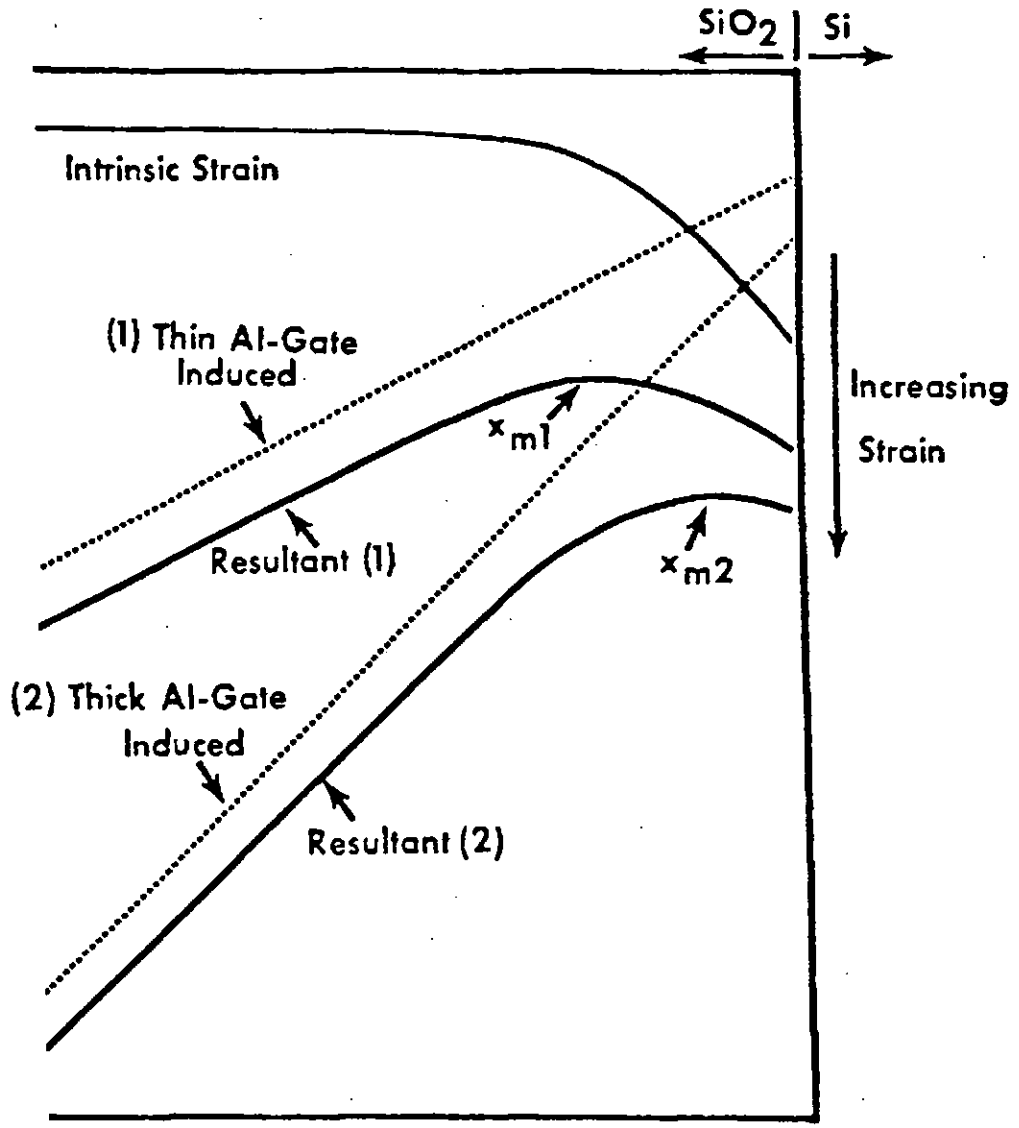


Figure 4.20 Model for the effect of the thickness of the aluminium gate electrode on interfacial stress. Taken from ref. (201) - Zekeriya and Ma.

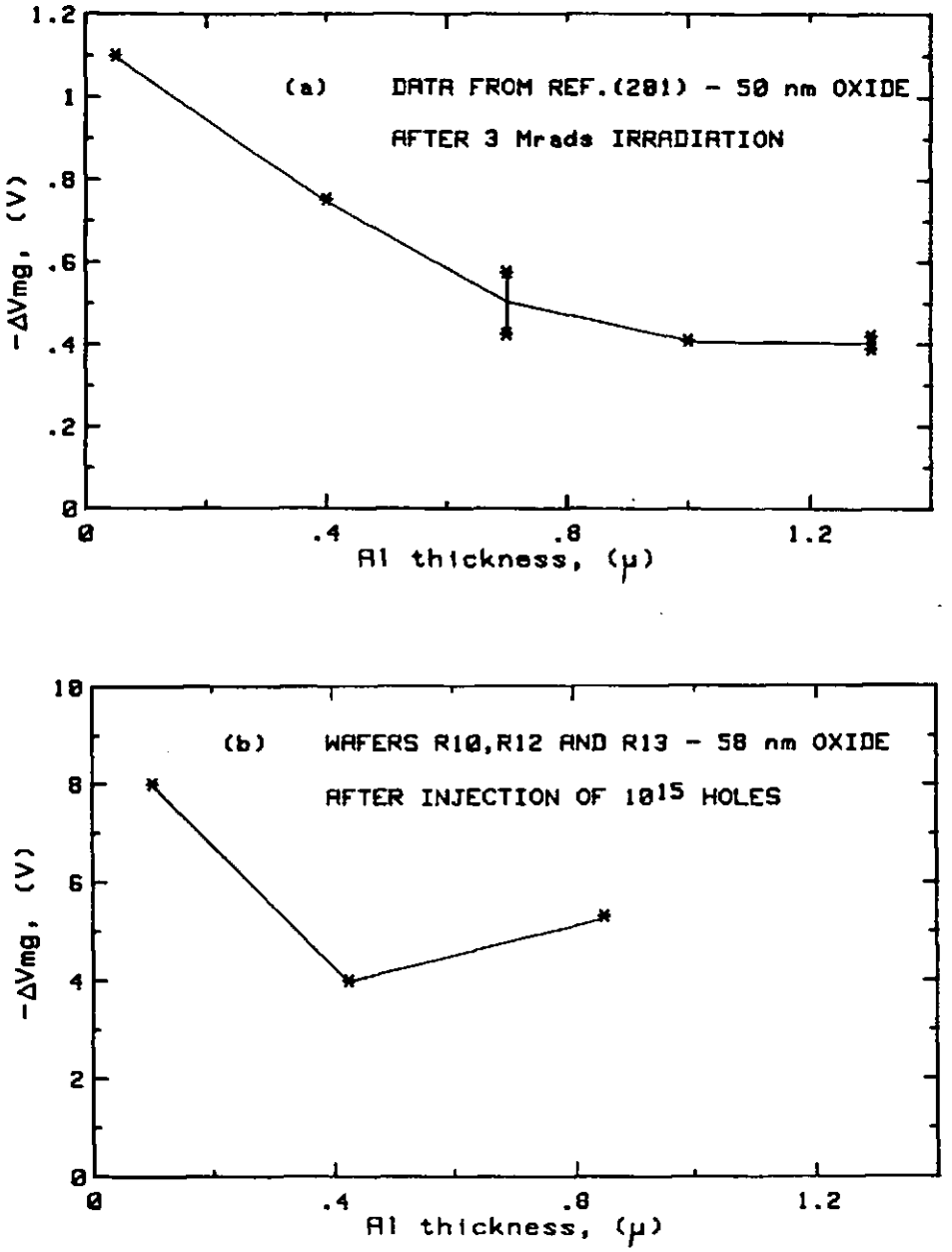


Figure 4.21 Variation of ΔV_{mg} with aluminium gate thickness (a) after irradiation (ref. (201)) and (b) after avalanche hole injection (present work - flash evaporated metal).

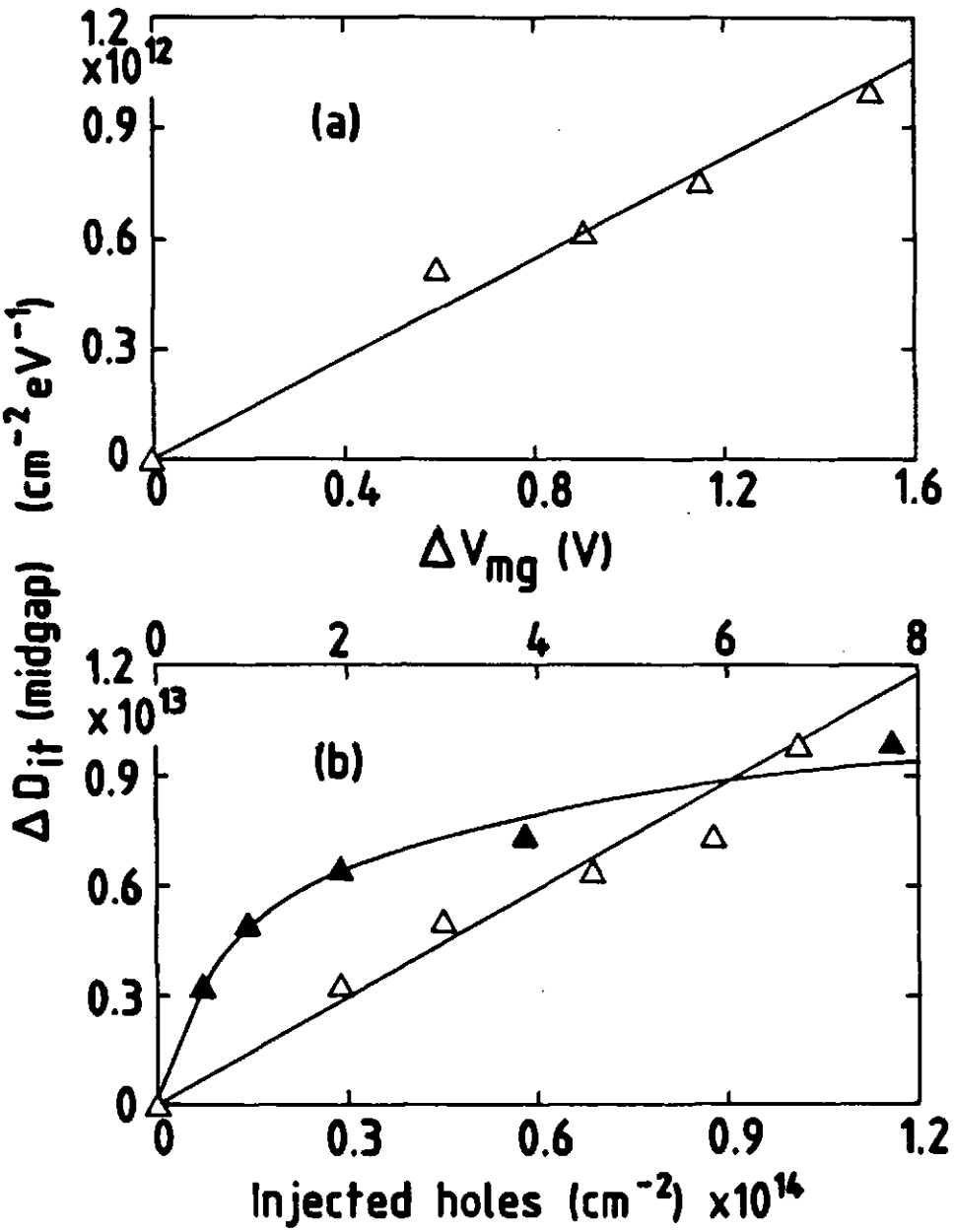


Figure 4.22 (a) Growth of ΔD_{it}(mg) as a function of ΔV_{mg} during negative BTS on wafer MN6 cap.F (b) Growth of ΔD_{it}(mg) during avalanche hole injection on wafer ST2, cap.F (△) as a function of ΔV_{mg} and (▲) as a function of injected charge

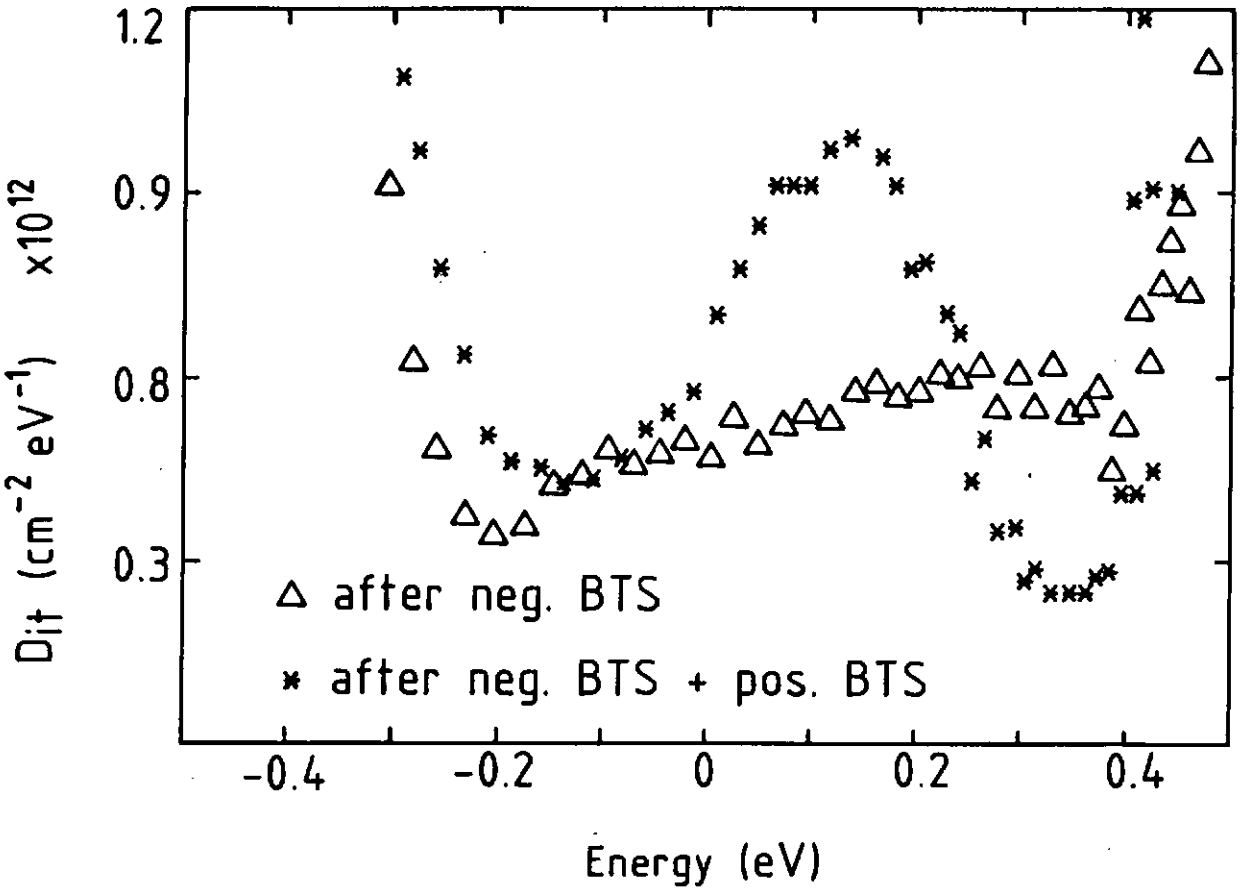


Figure 4.23 Interface state generation over the silicon bandgap following BTS on wafer MN6, cap.F. (Δ) After -2MV/cm at 250°C for 15 min. (*) After -2MV/cm at 250°C for 15 min and +2MV/cm at 250°C for 15 min.

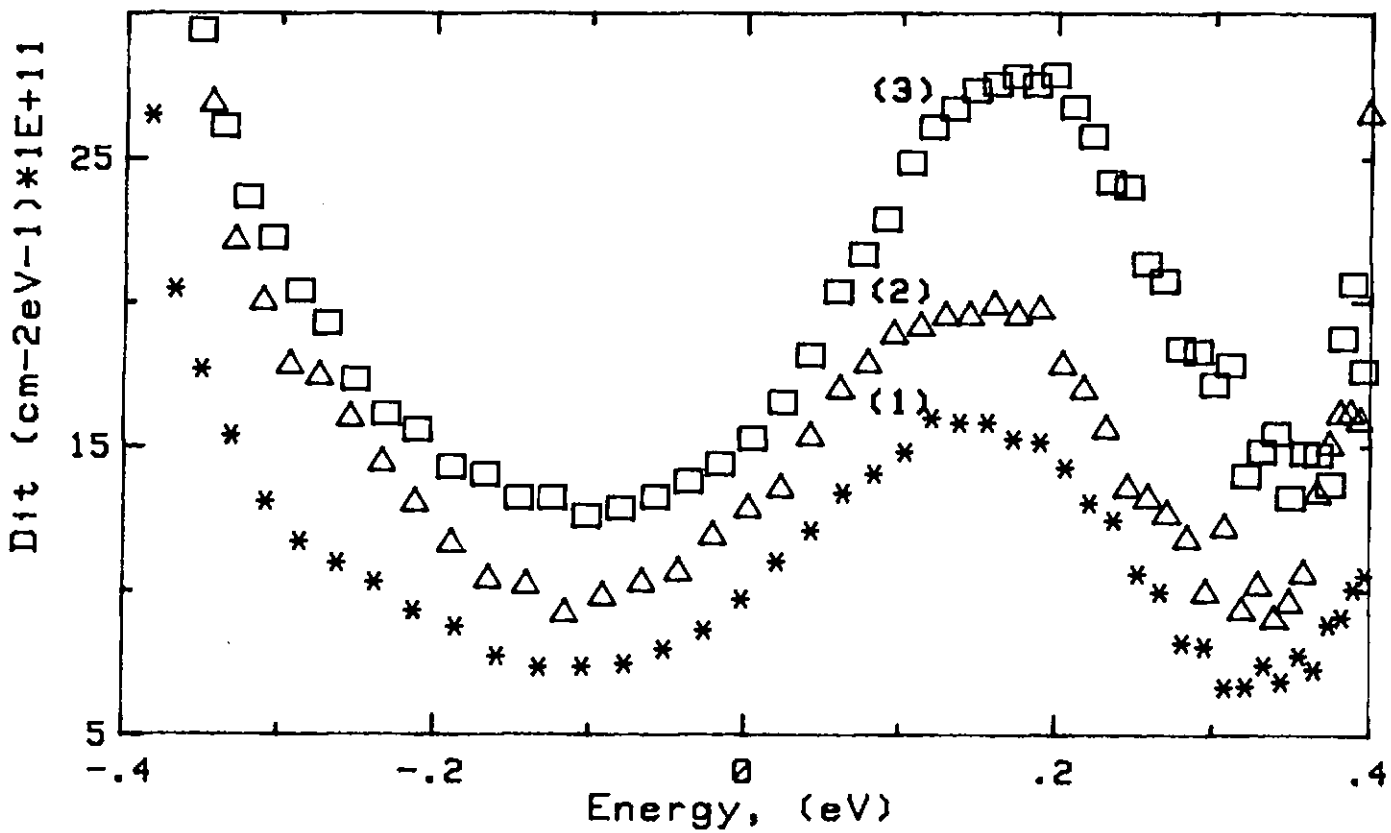


Figure 4.24 Interface state density distributions after negative BTS at 250°C and 2MV/cm on wafer ST2, cap.F for various times (1) 1 hr; (2) 4½ hr; (3) 11hr.20 min. followed by a short positive BTS

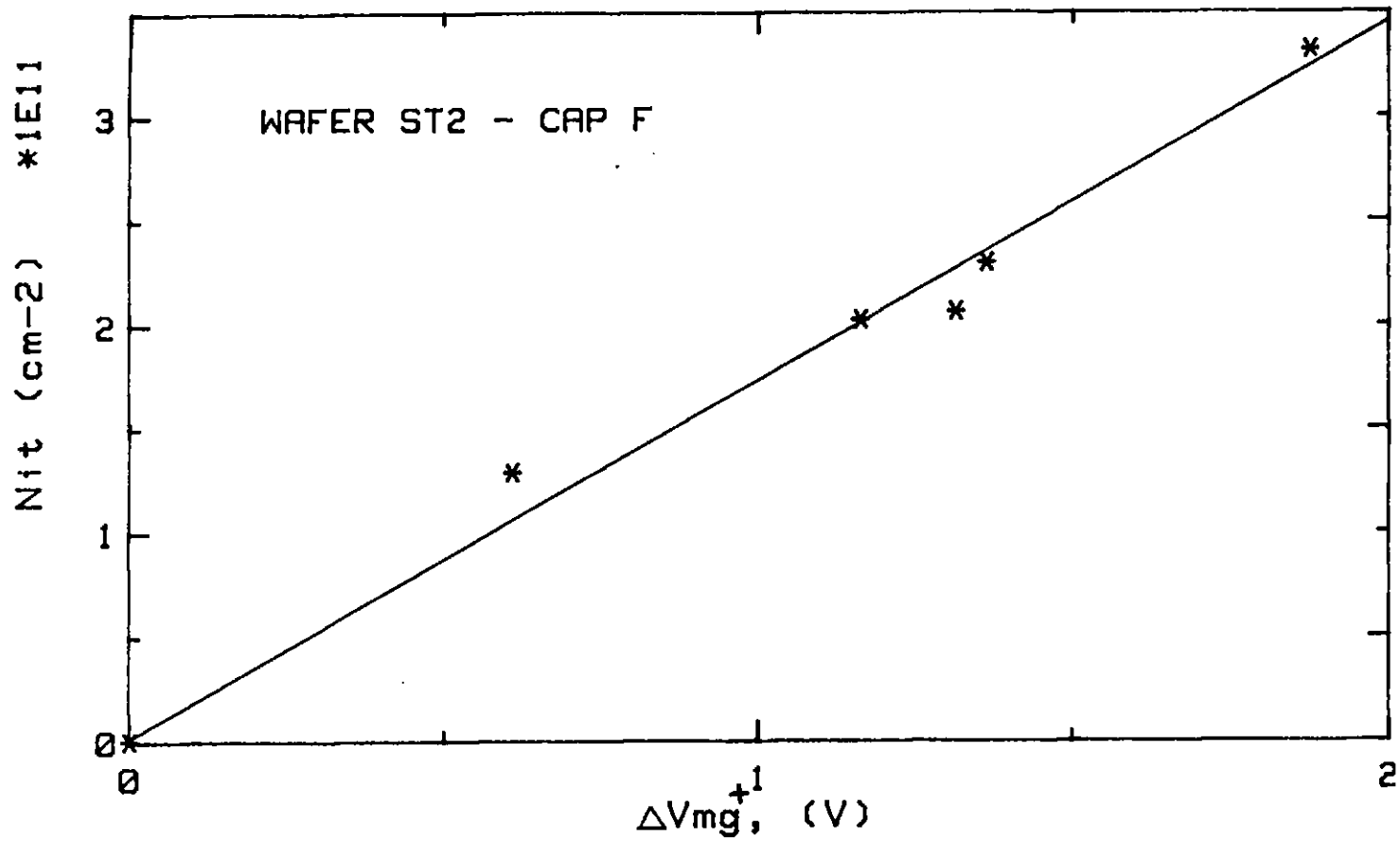


Figure 4.25 Area of the peak in the interface state distribution as a function of the change in V_{mg} on application of positive BTS after a previous negative BTS.

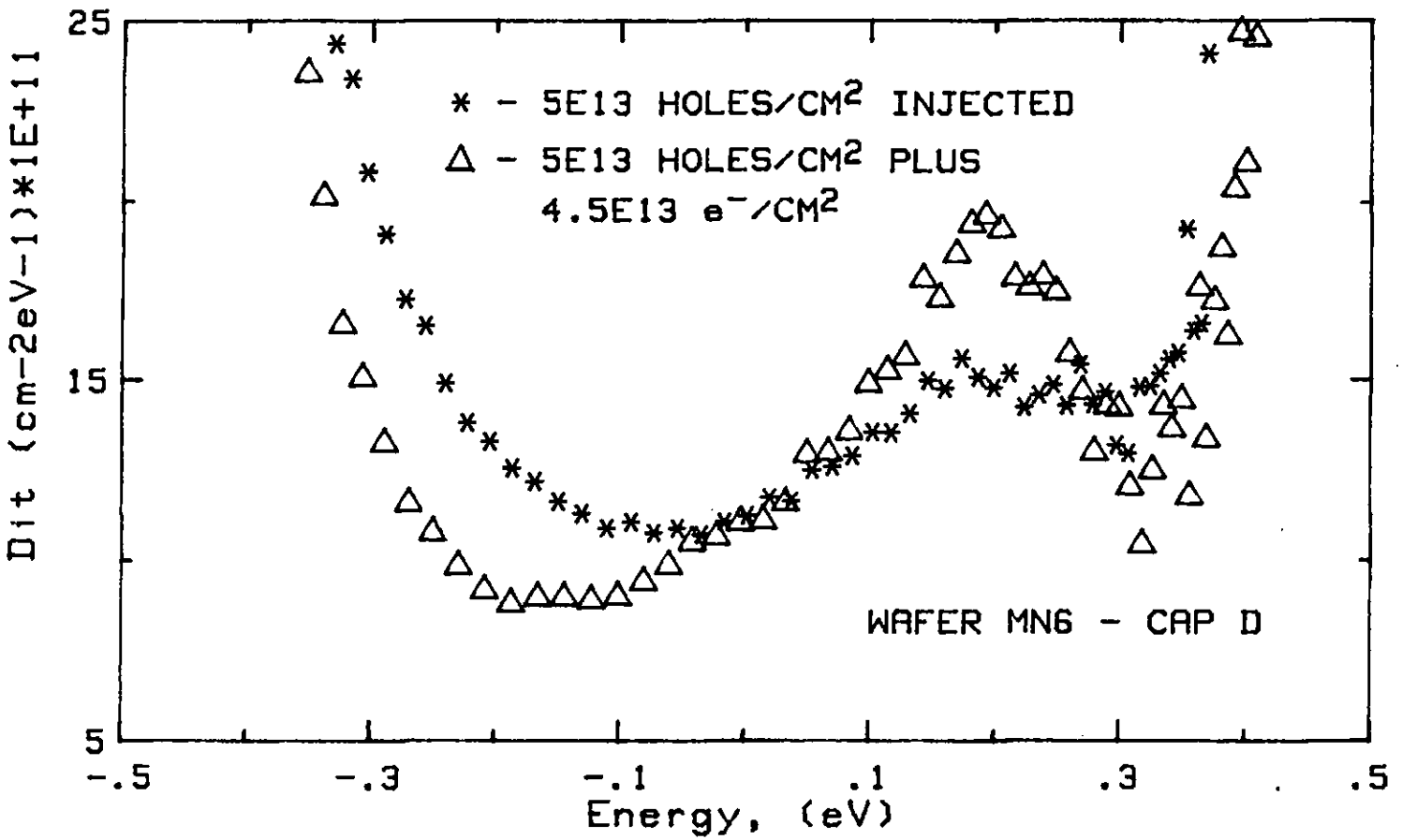


Figure 4.26 Interface state density distribution after avalanche hole injection (*) and after avalanche hole injection followed by photoinjection of electrons (Δ) on wafer MN6, cap.D.

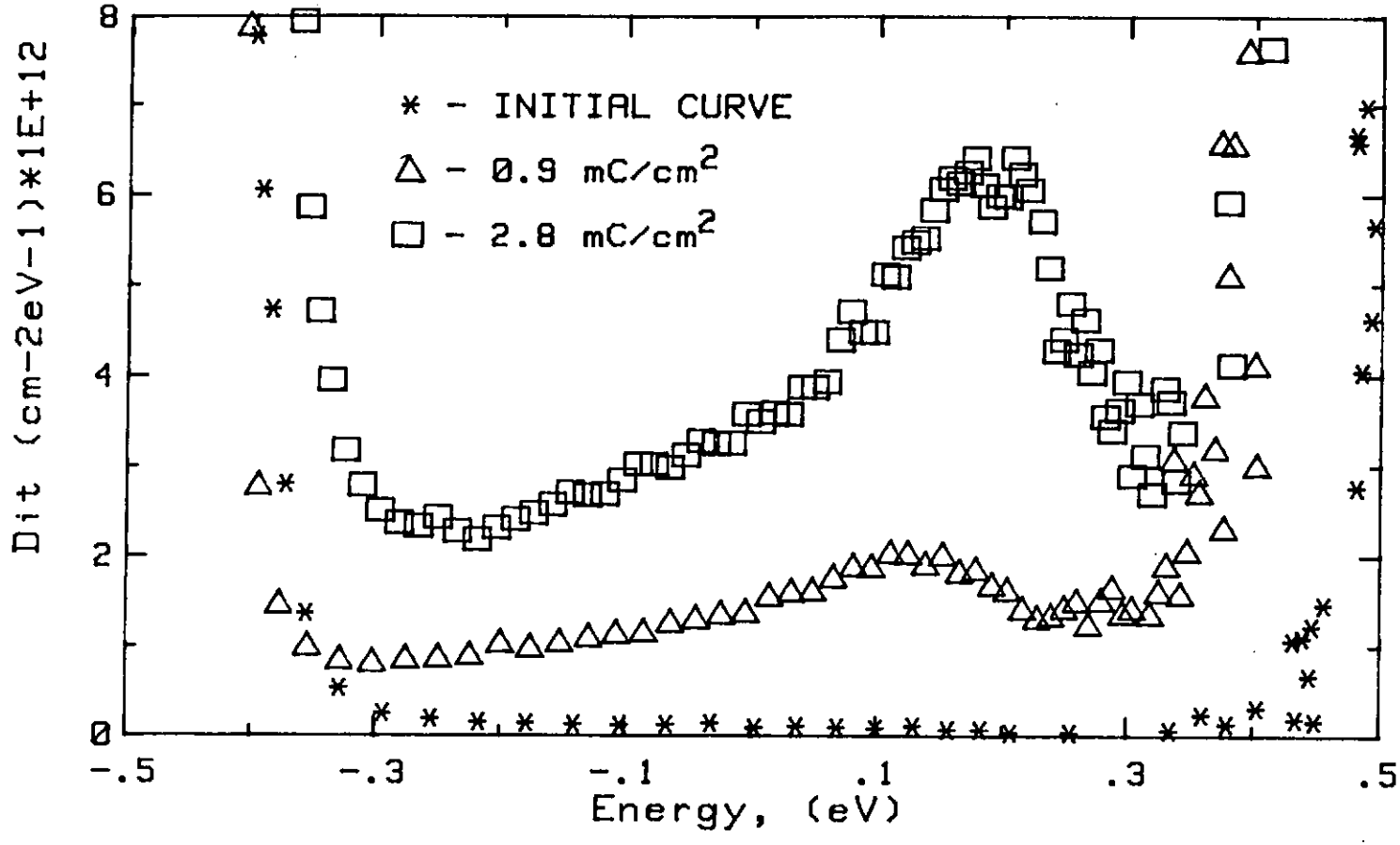


Figure 4.27 Interface state density distribution after injection of electrons by Fowler-Nordheim tunneling into wafer MN6 cap.D. The peak height increases with injected charge.

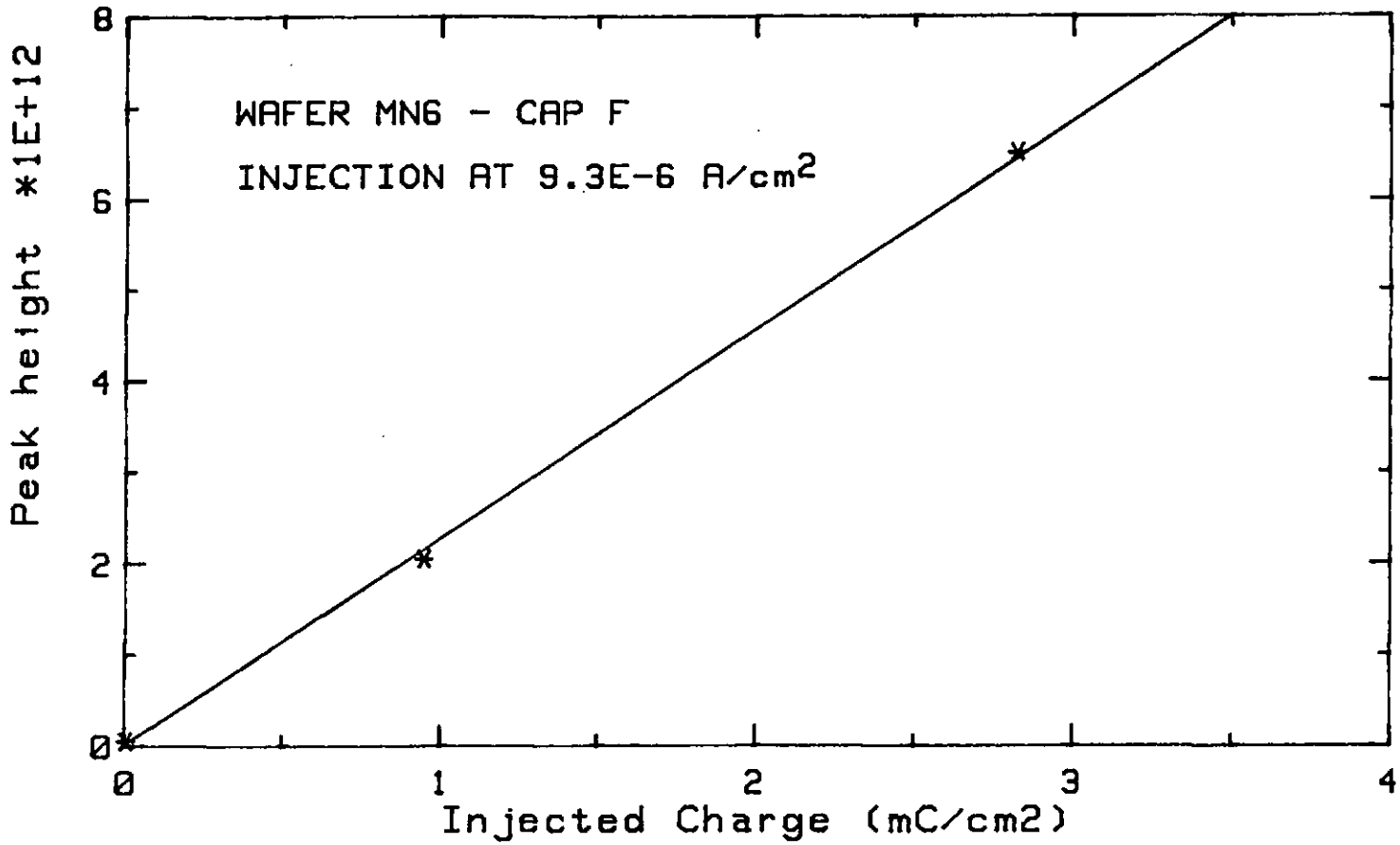


Figure 4.28 Variation of the height of the peak in the interface state distribution with charge injected by Fowler-Nordheim tunneling. (wafer MN6 - cap.F)

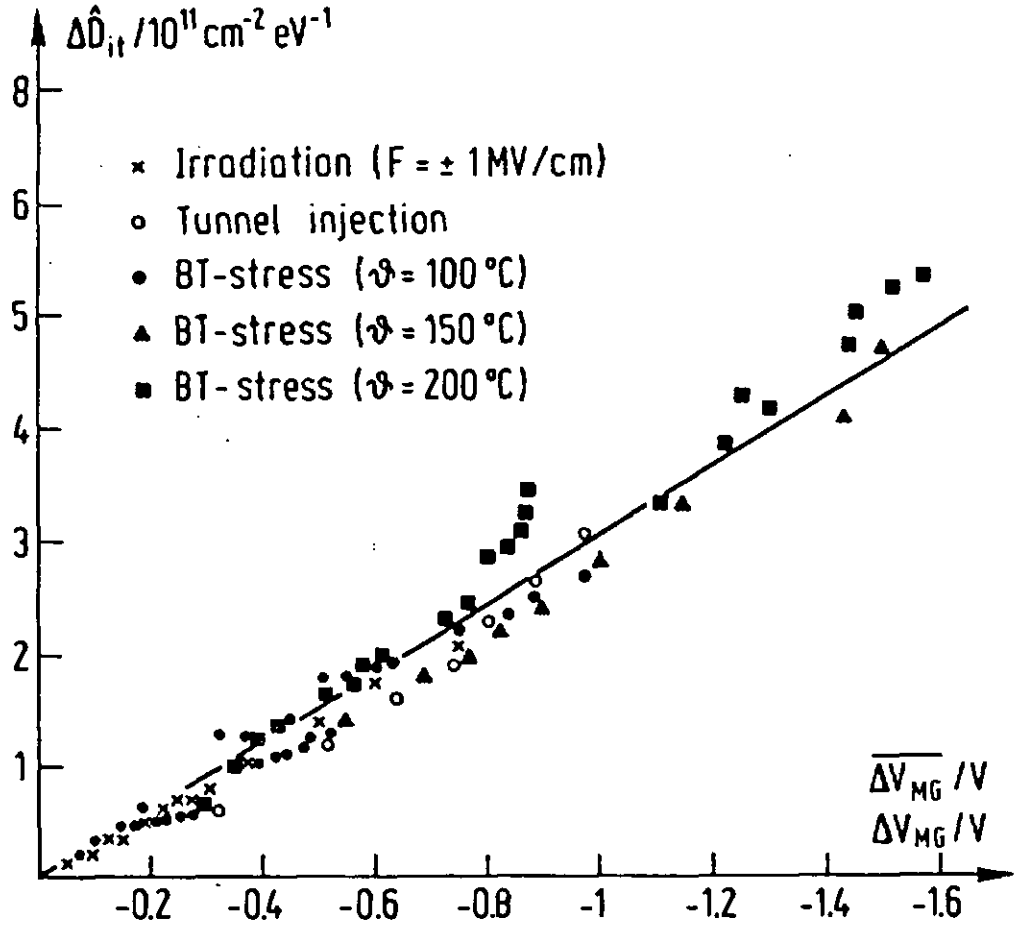


Figure 4.29 Variation of the height of the peak in the interface state distribution with ΔV_{mg} after irradiation and various electrical stresses. (Taken from ref. (48)).

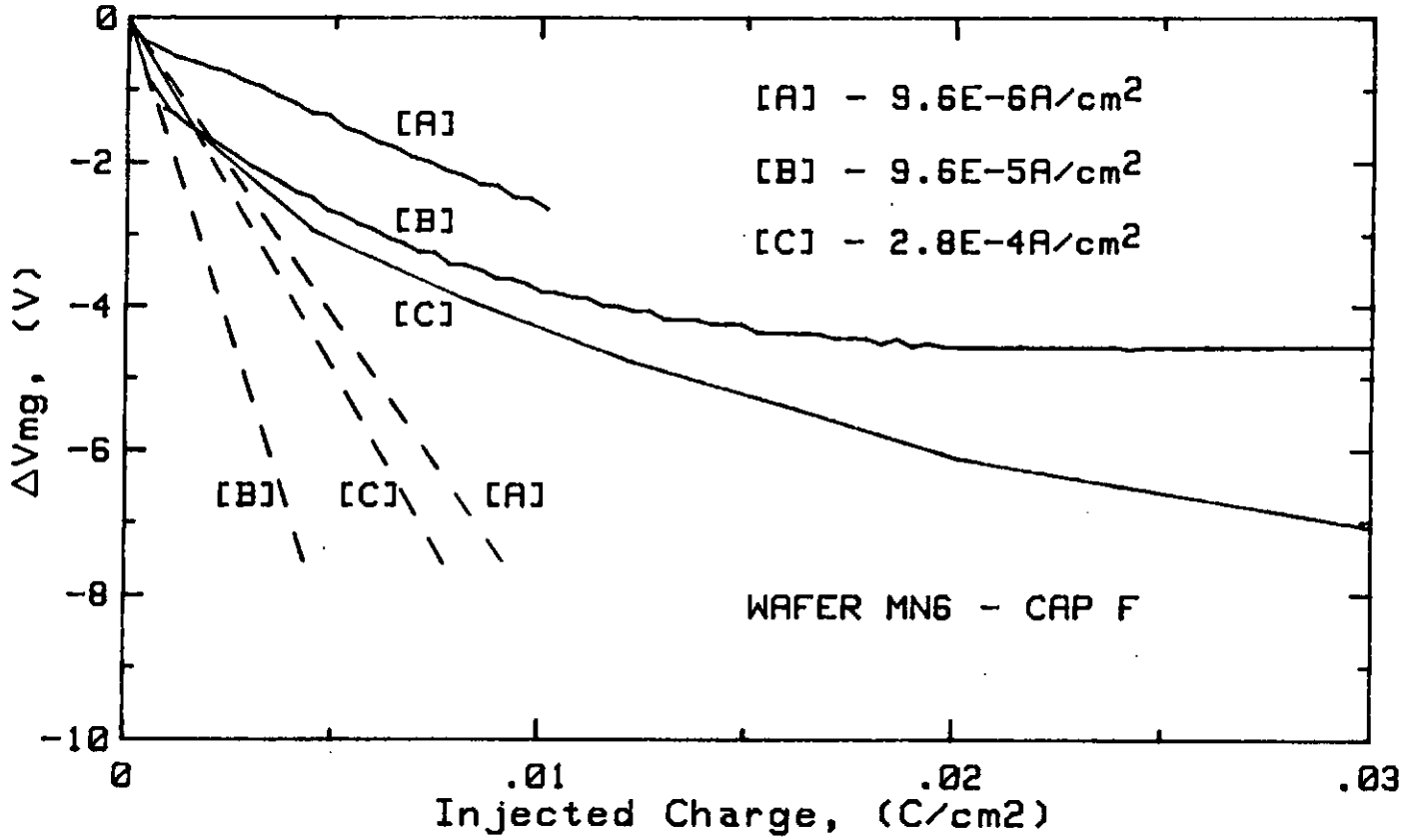


Figure 4.30 ΔV_{mg} as a function of the charge injected by F-N tunneling at various current densities. (Wafer MN6-cap.F). The initial slope of the curves measured varies with the injection current density.

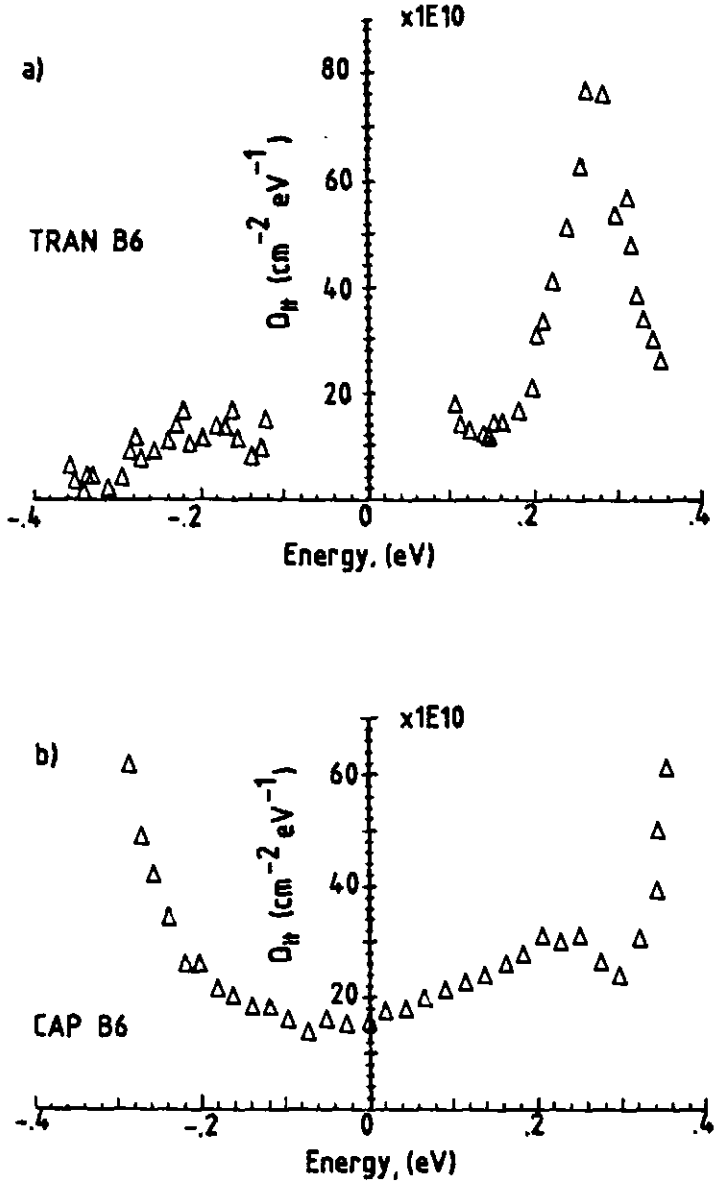


Figure 4.31 Interface state density distributions from devices on wafer B6 (a) by charge pumping (transistor) and (b) by the C-V method (capacitor). The gate insulator of both devices had been previously subjected to 9MV/cm for 120s followed by -2MV/cm at 250°C for ½ hr and +2MV/cm at 250°C for 15 min.

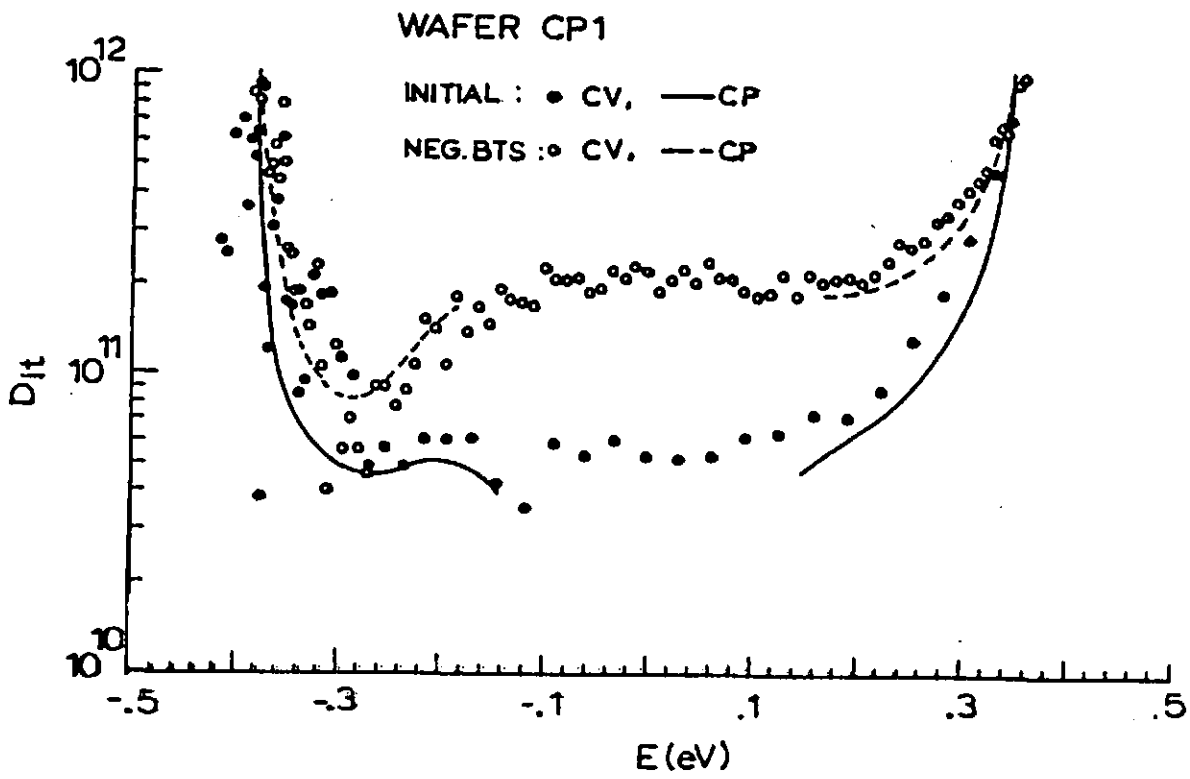


Figure 4.32 Interface state density distributions obtained by charge pumping (transistors) and by the C-V method (capacitors) on devices from wafer CP1 before and after negative BTS.

CHAPTER V:

CONCLUSIONS AND SUGGESTIONS FOR FURTHER WORK

Two major aspects of the electrical behaviour of the MOS capacitor system have been studied. Firstly, dielectric breakdown and the charge build-up and defect creation which immediately precede it were investigated. Secondly, the instability which is often seen during operation of devices under negative bias was examined by means of bias-temperature stress. The latter degradation occurs well before the breakdown condition is reached.

From a combination of photo I-V, avalanche electron injection and high field stress measurements it was demonstrated that both positive and negative charge are generated during high field stressing. The locations of this charge and of the electron traps created during the stress were established. The following picture resulted for the distribution of defects and charges in the oxide immediately prior to breakdown.

1) Positive charge resides in slow-states created at the $Si-SiO_2$ interface (within 10 Å of the silicon) after stress at either polarity.

2) Electron traps are created at the anode but under negative stress only a few of the energetically deep traps close to the interface (<30 Å) are filled.

3) During positive high field stress, electron traps near to the $Si-SiO_2$ interface are filled. In this case electrons must tunnel to the oxide conduction band to leave the oxide so the trap depth is the most important factor. Under negative stress they must only tunnel to the silicon so that the distance from the interface determines trap population.

4) No bulk charge was seen following high field stress under any injection conditions. All charge build-up and defect creation was within about 50Å of the interfaces.

5) With aluminium gate electrodes, positive charge is generated at the gate under positive high field stress.

The above observations point to the importance of the interfaces in understanding defect creation in the insulator. In particular, damage seems to be generated at the anode where the injected electrons attain their highest energy. Positively charged slow states were always observed at the $Si-SiO_2$ interface, however, regardless of whether this was the anode or the cathode.

Q_{bd}, the total charge which can be injected prior to breakdown was also investigated and was shown to be dependent on the current injection conditions. In particular, it varies linearly with the inverse of the applied field. Since during constant current injection on MOS capacitors, the current is related to the applied field by the F-N tunneling equation, Q_{bd} varies exponentially with the current density. Q_{bd} at positive polarity is increased by interspersing periods at lower positive or negative voltages, even when the negative stress results in current injection. The more negative (or less positive) the interrupting stress and the more frequent the interruption, the greater the increase in Q_{bd}. This suggests that relaxation of the voltage is an important factor in delaying breakdown. The gate electrode material and the measurement temperature have also been shown to affect Q_{bd}. Polysilicon gate capacitors show a much larger Q_{bd} than similariy processed aluminium gate devices, while increased temperature dramatically decreases Q_{bd}.

From the above, both current and field are obviously important in determining the point of breakdown. As the energy of the injected electrons increases (increasing field), Q_{bd} decreases but for

uninterrupted stress at a given applied field, Q_{bd} is constant. This implies that a certain amount of damage must be done to achieve the breakdown condition. Since evidence was found for defect creation at the anode, the two most likely breakdown models appearing in the literature would seem to be the gas discharge model of Budenstein and Wolters (31,32,33) and the anode hole injection model (144,164). The first of these depends on the trapping of electrons to precipitate breakdown and the latter on hole trapping. Further experiments are still required, therefore, to decide between these two main classes of breakdown model. Some indication of whether either electron trap generation or positive charge build-up is responsible for breakdown could perhaps be obtained from measurement of the rate of generation of these defects and of their density immediately prior to breakdown, for a number of different samples. If the density of either the positive charge or the electron traps does not saturate with time or if it always reaches a certain level before breakdown, this species would be a likely cause of breakdown.

Another experiment which remains to be conducted is the measurement of Q_{bd} at negative gate polarity for the same duty cycles as has been done for positive polarity. Similar results are expected, although a different injecting electrode may well cause a difference in the absolute magnitude of Q_{bd}. It would also be interesting to further investigate the temperature dependence of Q_{bd}, measuring enough samples at each temperature to carry out statistical analysis of the data.

The Weibull distribution was found to provide a useful statistical model to describe the observed breakdown distributions in both wearout and dielectric strength tests. Since the same values of the parameters a and b in the probability function were found for both measurements, the implication is that the mechanism of breakdown is the same in both cases. As a is less than one, the probability of breakdown occurring decreases with time. Electrical stressing therefore increases the reliability of the devices which survive the

test. This may be because under constant voltage conditions the injection current falls with time, as has been suggested (123). Measurement of the failure distribution and evaluation of a and b under constant current conditions could show whether this is the case. This might also prove interesting from another point of view. The Weibull distribution, as a statistical description of the breakdown process, was arrived at theoretically by Hill and Dissado (152-155) from a physical model which does not take into account any change in the applied field due to charge trapping in the oxide. Constant current stress should in fact provide a closer approximation to constant field conditions than constant voltage stress. It would be interesting, therefore, to see whether the same values of a and b are found for constant current wearout as for dielectric strength measurements.

The second major topic studied here was the negative bias instability. It was found from avalanche hole injection experiments on samples having undergone negative BTS and on unstressed samples, that the positive charge generated during negative BTS is due to the filling of intrinsic hole traps. These are located within 50 nm of the $Si-SiO_2$ interface. Since no new hole traps are created during BTS the number of intrinsic hole traps in an as-processed oxide appears to be an extremely important parameter. It is a strong indicator of the degradation likely to occur during operation of the device into which the oxide is incorporated. The fact that hole trapping and de-trapping have both been shown to play a key role in interface state generation increases the importance of these intrinsic hole traps still further.

Although the positive charge is now known to result from hole trapping, two questions remain to be answered. Neither the mechanism of hole trap filling during negative BTS nor the physical nature of a hole trap have been clearly established. The fact that no current flow was observed during negative BTS suggests that some sort of trap hopping or thermally activated tunneling process is occurring by

which holes may enter the insulator. Almost certainly the process is different from that for positive charge generation during current injection under avalanche or F-N tunneling conditions. A modified floating gate technique could perhaps be used to measure more accurately the current during BTS in order to establish whether this is actually zero or merely very low. This method allows currents down to about 10^{-17} A to be measured (203).

The kinetics of the negative bias instability were investigated using the midgap voltage as an indicator of the change in oxide charge, since it was found empirically that this represented a neutral point for interface state charge. $\Delta V_{mg} \propto \ln t$ was found always to give a reasonably good fit to the data. A clear dependence of ΔV_{mg} on $E_{ox}^{1.5}$ was observed. Unfortunately, it is difficult to deduce any particular model for hole trapping from these kinetic data. The value of the activation energy measured (around 0.3 eV) was perhaps the most useful result from this point of view. It is close to the value expected for the breaking of a hydrogen bond and compatible with the trapping of holes at $O_3{\equiv}Si\cdot \quad \cdot Si{\equiv}O_3$ or $O_3{\equiv}Si\cdot \quad \cdot O\text{-}Si{\equiv}O_3$ sites coordinated by water as proposed in ref. (143).

The role of interfacial stress and of water in generating hole traps remains to be clarified. The dependence of the trap density on the thickness of flash-evaporated aluminium gates suggests that stress does play a role. Correlation of results from electrical measurements and from mechanical stress measurements are required to confirm this, however. The latter may be performed on both a macroscopic scale, using optical techniques, and on a microscopic scale, by means of Raman spectroscopy. Other effects of the electrode material also remain to be further investigated. For example, polysilicon gate devices are known from avalanche hole injection measurements to have a large number of hole traps. However, these are not readily filled under BTS conditions. More detailed investigation of the density and location of hole traps as a function of the electrode material and its method of deposition are needed to establish the reason for this.

Water certainly enhances the negative bias instability when it is allowed to diffuse into the oxide after growth (60,61). Wafers with different water content and with the water incorporated during and after growth could be used to obtain a more precise picture of the relationship between water-related species and hole traps. Infra-red spectroscopy provides a potential method of measuring the density of Si-H and Si-OH bonds in the differently processed wafers.

During BTS, interface state generation has been shown to be proportional to ΔV_{mg}, i.e. to the number of trapped holes. No evidence was found for the previously reported one-to-one ratio of these defects, however (49). The ratio of interface states generated to holes trapped was rather high for the samples measured in this work when compared with similar results in the literature. The sensitivity of samples to interface state generation a result of hole trapping would appear to depend on the processing, therefore. The linear relationship of interface state generation to hole trapping suggests that both of these may be due to the same defect. This could give rise to an interface state when located at the interface or to a hole trap when situated further into the oxide. The amphoteric site ≡Si• •O-Si≡ seems a likely candidate for this defect since ≡Si• is known to be responsible for the U-shaped distribution found in as-processed oxides and since interface state generation accompanying hole trapping also occurs across the whole band gap.

Positive BTS after negative BTS led to the appearance of a peak in the interface state distribution at about 0.2 eV above midgap. This could be due to either a redistribution of existing states to this specific energy level or to the generation of new states. A redistribution in energy seems the most likely explanation for the reversible removal and regeneration of the peak by negative BTS and positive BTS, respectively and also for the removal of states from other parts of the band gap when the peak is created.

CONCLUSIONS

Using the charge pumping technique it has been established that the peak at 0.2 eV above midgap is due to interface states and not to lateral non-uniformities in the surface potential. A peak at this energy has been seen following avalanche hole injection and subsequent photo I-V, high field stress and irradiation. All these techniques supply both electrons and holes to the oxide. It seems quite general, therefore, that hole trapping followed by either neutralisation or de-trapping of the holes generates this peak. Although the interface degradation is qualitatively similar for these different techniques, it is difficult to compare the effect of hole trapping on the interface quantitatively. Other processes such as electron injection and trapping may also cause oxide degradation to a varying degree in the different measurements.

It would be interesting to carry out further charge pumping measurements on samples likely to show a peak in the interface state density. The charge pumping technique was found to give better resolution of the peak than C-V methods and thus to provide a potentially more accurate means of examining the conditions which generate it. For example, although it seems that the peak usually accompanies de-trapping or neutralisation of trapped holes, the peak was <u>not</u> seen following channel hot hole injection and subsequent channel hot electron injection (204). This led these authors to suggest that excitons are responsible for its generation i.e. trapped holes must first be de-trapped and it is the process of recombination with incoming electrons and the accompanying release of energy which results in interface state generation. More experiments are needed to establish whether this is the case or if it is merely that the holes and electrons are not in the same physical location in the case of inhomogeneous injection from the channel of a MOSFET.

Breakdown and charge trapping in the MOS system have been studied here from an electrical point of view. Knowledge of these electrical properties is, of course, of primary importance for device fabrication. However, perhaps the most important direction for future

work in this field is a more thorough physical characterisation of the oxide interface region to correlate structure and electrical properties. This is certainly essential to the formulation of accurate models both for the breakdown mechanism and for hole trapping and the accompanying interface state generation.

REFERENCES:

1) G.Moore,
IEEE Spectrum, 16, 30 (1979).

2) J.Yamada, T.Mano, J.Inoue, S.Nakajima and T.Matsuda,
IEEE J.Solid State Circuits, 19, 627 (1984).

3) R.Hori, K.Hoh, J.Etoh, S.Asai, N.Hashimoto, K.Yagi and H.Sunami,
IEEE J.Solid State Circuits, 19, 634 (1984).

4) "Looking ahead to the year 2000: Technology",
Electronics, 53, 530 (1980).

5) C.M.Osburn and D.W.Ormond,
J.Electrochem.Soc., 119, 597 (1972).

6) T.H.DiStefano and M.Shatzkes,
J.Vac.Sci.Tech., 13, 50 (1976).

7) A.C.Adams, T.E.Smith and C.C.Chang,
J.Electrochem.Soc., 127, 1787 (1980).

8) E.Harari,
J.Appl.Phys., 49, 2478 (1978).

9) T.Ogata and K.Okamoto,
Rep.Univ.Electro-Commun. (Japan), 28, 207 (1978).

10) P.V.Dressendorfer, S.K.Lai, R.C.Barker and T.P.Ma,
Appl.Phys.Lett., 36, 850 (1980).

11) E.H.Nicollian and J.R.Brews,
"MOS Physics and Technology", Wiley (New York) (1982).

12) M.Lenzlinger and E.H.Snow,
J.Appl.Phys., 40, 278 (1969).

13) C.M.Osburn and N.J.Chou,
J.Electrochem.Soc., 120, 1377 (1973).

14) D.L.Crook,
Proc.IEEE.Int.Rel.Symp., p1, (1979).

15) C.Fritzsche,
Z.Angew.Phys., 24, 49 (1967).

16) N.J.Chou and J.M.Eldridge,
J.Electrochem.Soc., 117, 1287 (1970).

17) M.W.Hillen, R.F.DeKeersmaecker, M.M.Heyns, S.K.Haywood and
I.S.Darakchiev,
IEEE.Trans.Elec.Insulation EI-19, 245 (1984).

18) D.J.DiMaria, T.N.Theis, J.R.Kirtley, F.L.Pesavento and
S.D.Brorson,
J.Appl.Phys., 58, 1302 (1984).

19) C.M.Osburn and E.Bassous,
J.Electrochem.Soc., 122, 89 (1975).

20) Z.A.Weinberg, W.C.Johnson and M.A.Lampert,
J.Appl.Phys., 47, 248 (1976).

21) M.W.Hillen, R.F.DeKeersmaecker, M.M.Heyns, S.K.Haywood and I.S.Darakchiev,
"Insulating Films on Semiconductors" p274,
Eds. J.F.Verweij and O.R.Wolters, North Holland (1983).

22) P.F.Schmidt,
J.Electrochem.Soc., 130, 196 (1983).

23) C.M.Osburn and D.W.Ormond,
J.Electrochem.Soc., 119, 591 (1972).

24) C.M.Osburn and E.J.Weitzman,
J.Electrochem.Soc., 119, 603 (1972).

25) T.H.DiStefano and M.Shatzkes,
Appl.Phys.Lett., 25, 684 (1974).

26) P.Solomon and N.Klein,
Solid State Comm., 17, 1397 (1975).

27) M.Shatzkes and M.Av-Ron,
J.Appl.Phys., 47, 3192 (1976).

28) T.H.DiStefano and M.Shatzkes,
J.Vac.Sci.Tech., 13, 50 (1976).

29) N.Klein and P.Solomon,
J.Appl.Phys., 47, 4364 (1976).

30) E.Harari,
Appl.Phys.Lett., 30, 601 (1979).

31) P.Budenstein,
IEEE Trans.Elec.Insulation EI-15, 225 (1980).

32) D.R.Wolters, in "Insulating Films on Semiconductors" p180,
Eds. M.Schulz and G.Pensl, Springer-Verlag (1981).

33) D.R.Wolters, in "The Physics of MOS Insulators" p349, (1980)
Ed. G.Lucovsky, S.Pantelides and F.L.Galeener.

34) E.H.Nicollian and C.N.Berglund,
J.Appl.Phys., 41, 3052 (1970).

35) D.J.DiMaria,
J.Appl.Phys. 47, 4073 (1976).

36) D.R.Wolters and J.J.Van der Schoot,
Philips J.Research, to be published.

37) J.J.VanderSchoot and D.R.Wolters,
"Insulating Films on Semiconductors", p270,
Eds. J.F.Verweij and D.R.Wolters, North Holland (1983).

38) E.S.Anolick and G.R.Nelson,
Proc.IEEE.Int.Rel.Phys.Symp., p8, (1979).

39) A.Berman,
Proc.IEEE.Int.Rel.Phys.Symp., p208, (1981).

40) B.E.Deal,
J.Electrochem.Soc., 127, 979 (1980).

41) E.H.Snow, A.S.Grove, B.E.Deal and C.T.Sah,
J.Appl.Phys., 36, 1664 (1965).

42) B.E.Deal,
J.Electrochem.Soc., 121, 198C (1974).

43) K.L.Wang,
"Semiconductor Silicon 1977", p404,
Eds. H.R.Huff and E.Sirtl, The Electrochemical Society, 1977.

44) L.Risch, E.Pammer and K.Friedrich,
"Insulating Films on Semiconductors (1979)", p62,
Eds. G.G.Roberts and M.J.Morant, Inst.Phys.Conf.Ser., 50 (1980).

45) P.J.Caplan, E.H.Poindexter, B.E.Deal and R.R.Razouk,
J.Appl.Phys., 50, 5847 (1979).

46) S.K.Lai and D.R.Young,
J.Appl.Phys., 52, 6231 (1981).

47) S.K.Lai,
J.Appl.Phys., 54, 2540 (1983).

48) G.Haller, M.Knoll, D.Braunig, F.Wulf and W.R.Fahrner,
J.Appl.Phys., 56, 1844 (1984).

49) K.O.Jeppson and C.M.Svensson,
J.Appl.Phys., 48, 2004 (1977).

50) G.A.Scoggan and T.P.Ma,
J.Appl.Phys., 48, 294 (1977).

51) P.S.Winokur, H.E.Boesch, J.M.McGarrity and F.B.McClean,
J.Appl.Phys., 50, 3492 (1979).

52) P.S.Winokur and H.E.Boesch,
IEEE Trans.Nuc.Sci. NS-27, 1647 (1980).

53) M.M.Heyns and R.F.DeKeersmaecker,
Presented at 6th Symposium SSDT (ESSDERC),
Toulouse, 14th-16th Sept. (1981).

54) R.A.Gdula,
J.Electrochem.Soc., 123, 42 (1976).

55) D.R.Young, E.A.Irene, D.J.DiMaria, R.F.DeKeersmaecker and H.Z.Massoud,
J.Appl.Phys., 50, 6366 (1979).

56) M.M.Heyns, private communication.

57) R.R.Razouk and B.E.Deal,
J.Electrochem.Soc., 126, 1573 (1979).

58) B.E.Deal, M.Sklar, A.S.Grove and E.H.Snow,
J.Electrochem.Soc., 114, 266 (1967).

59) E.H.Nicollian,
J.Vac.Sci.Tech., 14, 112 (1977).

60) M.Noyori and Y.Nakata,
J.Electrochem.Soc., 131, 1109 (1984).

61) M.Noyori, Y.Nakata and T.Shiragasawa,
IEEE.Trans.Electron.Devices ED-30, 1305 (1983).

62) D.J.Breed and R.P.Kramer,
Solid State Electron., 19, 897 (1976).

63) N.Shiono, O.Nakajima and C.Hashimoto,
J.Electrochem.Soc., 130, 138 (1983).

64) A.Goetzberger, A.D.Lopez and R.J.Strain,
J.Electrochem.Soc., 120, 90 (1973).

65) S.R.Hofstein,
Solid State Electron., 10, 657 (1967).

66) D.J.Breed,
Solid State Electron., 17, 1229 (1974).

67) D.J.Breed,
Appl.Phys.Lett., 26., 116 (1975).

68) J.Bennett and L.M.Roth,
J.Phys.Chem.Solids, 32, 1251 (1971).

69) R.F.DeKeersmaecker, M.W.Hillen, M.M.Heyns, S.K.Haywood and I.S.Darakchiev,
Presented at 2nd Int'l VLSI Symp., 165th meeting of the Electrochemical Society, Cincinnati, May 6-11th (1984).

70) P.A.Heiman,
IEEE.Trans.Electron.Devices ED-30, 1366 (1983).

71) P.Solomon,
J.Appl.Phys., 48, 3843 (1977).

72) S.Holland, I-C.Chen,T.P.Ma and C.Hu,
IEEE Electron.Dev.Lett. EDL-5, 302 (1984).

73) I-C.Chen, S.E.Holland and C.Hu,
IEEE.Trans.Electron.Devices ED-32, 413, (1985).

74) K.Nagai, Y.Hayashi and Y.Tarui,
Jap.J.Appl.Phys., 14, 1539 (1978).

75) J.M.Aitken and D.R.Young,
IEEE Trans.Nuc.Sci NS-24, 2128 (1977).

76) R.Williams,
Phys.Rev., 140, A569 (1969).

77) R.J.Powell and C.N.Berglund,
J.Appl.Phys., 42, 4390 (1971).

78) R.E.Burgess and H.Kroemer,
Phys.Rev., 90, 515 (1953).

79) A.Rush and C.Bulucea,
IEEE.Trans.Electron.Devices ED-26, 201 (1979).

80) T.H.Ning and H.N.Yu,
J.Appl.Phys., 45, 5375 (1974).

81) A.Bjorck,
"Least squares methods in physics and engineering",
CERN report 81-16 (1981).

82) L.Trombetta,
Ph.D Thesis, Lehigh University (1984).

83) S.M.Sze,
"Physics of Semiconductor Devices", Wiley (1981).

84) A.Harstein and Z.A.Weinberg,
J.Phys.C., 11, L469 (1978).

85) D.J.DiMaria, Z.A.Weinberg, J.M.Aitken and D.R.Young,
J.Electron.Mat., 6, 207 (1977).

86) M.Kuhn,
Solid State Electron., 13, 873 (1970).

87) A.D.Lopez,
Rev.Sci.Inst., 44, 200 (1972).

88) A.S.Grove, B.E.Deal, E.H.Snow and C.T.Sah,
Solid State Electron., 8, 145 (1965).

89) F.J.Feigl, D.R.Young. D.J.DiMaria, S.K.Lai and J.Calise,
J.Appl.Phys., 52, 5665 (1981).

90) C.N.Berglund,
IEEE.Trans.Electron.Devices ED-13, 701 (1966).

91) R.Castagne and A.Vapaille,
Surf.Sci. 28, 557 (1971).

92) J.S.Brugler and P.G.A.Jespers,
IEEE.Trans.Electron.Devices ED-16, 297 (1969).

93) G.Groeseneken, H.Maes, N.Beltran and R.F.DeKeersmaecker,
IEEE.Trans.Electron.Devices ED-31, 42 (1984).

94) J.G.Simmons and L.S.Wei,
Solid State Electron., 16, 53 (1973).

95) G.Groeseneken, H.E.Maes, N.Beltran and R.F.DeKeersmaecker,
"Insulating Films on Semiconductors", p153,
Eds. J.F.Verweij and D.R.Wolters, North Holland (1983).

96) N.Klein,
Thin Solid Films, 50, 223 (1978).

97) P.Solomon,
J.Vac.Sci.Tech., 14, 112, (1977).

98) N.Klein,
Thin Solid Films, 100, 335 91983).

99) W.Kern,
RCA Rev., 34, 655 (1975).

100) J.M.Keen,
Electron.Lett., 7, 432 (1971).

101) A.K.Zakzouk, R.A.Stuart and W.Eccleston,
J.Electrochem.Soc., 123, 1551 (1976).

102) R.Williams and M.H.Woods,
J.Appl.Phys., 43; 4142 (1972).

103) T.H.DiStefano,
Appl.Phys.Lett., 19, 280 (1971).

104) J.W.Philbrick and T.H.DiStefano,
Proc.13th.Ann.Rel.Phys.Conf., p159,
Las Vegas, April 1st. (1975).

105) M.Shatzkes, M.Av-Ron and R.Gdula,
IBM J.Res.Dev., 24, 469 (1980).

106) C.M.Osburn and M.Raider,
J.Electrochem.Soc., 120, 1369 (1973).

107) T.H.DiStefano,
J.Appl.Phys., 44, 527 (1973).

108) K.Yamabe and K.Taniguchi,
IEEE.Trans.Electron.Devices ED-32, 423 (1985).

109) S.P.Li and J.Maserjian,
IEEE.Trans,Electron.Devices ED-23, 525 (1976).

110) B.H.Vromen,
Appl.Phys.Lett., 27, 152 (1975).

111) I.Darakchiev,
M.Sc. Thesis, K.U. Leuven, Belgium (1984).

112) C.Hashimoto, S.Muramoto, N.Shiono and O.Nakajima,
J.Electrochem.Soc., 127, 129, (1980).

113) C.M.Osburn,
J.Electrochem.Soc., 121, 869 (1974).

114) M.Henzler,
Invited paper INFOS '85, Toulouse, April 16-18th (1985).

115) R.L.Meek and R.H.Braun,
J.Electrochem.Soc., 119, 1538 (1972).

116) S.I.Raider,
Appl.Phys.Lett., 46, 4833 (1975).

117) C.M.Osburn and D.W.Ormond,
J.Electrochem.Soc., 121, 1195 (1974).

118) F.L.Worthing,
J.Electrochem.Soc., 115, 89 (1968).

119) J.G.Price,
Proc.IEEE, 58, 1290 (1970).

120) F.Jensen and N.E.Petersen,
"Burn-in", Wiley, New York (1982).

121) W.Kuo and Y.Kuo,
Proc.IEEE, 71, 1257 (1983).

122) D.W.Ormond,
J.Electrochem.Soc., 126, 162 (1979).

123) D.Wolters and J.Verweij,
to be published in "Electrical Phenomena and Related Instabilities in the Si-SiO@2 Structure - Impact on Modern Integrated Devices.", Ed. G.Barbottin.

124) J.W.McPherson and D.Baglee,
J.Electrochem.Soc., 132, 1903 (1985).

125) Y.Hokari, T.Baba and N.Kawamura,
IEDM.Tech.Digest, p46, (1982).

126) D.A.Baglee,
Proc.22nd.IEEE.Int.Rel.Phys.Symp., (1984).

127) N.Klein,
Thin Solid Films, 7, 149 (1971).

128) W.W.Wagner,
Trans.AIEE, 41, 288 (1922).

129) B.K.Ridley,
J.Appl.Phys., 46, 978 (1975).

130) P.Budenstein and J.M.Lloyd,
Proc.Conf.Elect.Insulation and Dielectric Phenomena, p303,
Gaithersburg, Nov. 3-6th. (1975).

131) P.Budenstein, P.J.Hayes, J.L.Smith and W.B.Smith,
J.Vac.Sci.Tech., 6, 289 (1969).

132) B.Ricco, M.Ya.Azbel and M.Brodsky,
Phys.Rev.Lett., 51, 1795 (1983).

133) B.P.Rai, K.Singh, and R.Srivasttava,
Phys.Status Solidi, 36, 591 (1976).

134) J.J.O'Dwyer,
J.Appl.Phys., 40, 3887 (1969).

135) R.C.Hughes,
Phys.Rev.Lett., 55, 449 (1975).

136) R.C.Hughes,
Phys.Rev.B, 715, 2012 (1977).

137) D.K.Ferry,
Solid State Comm., 18, 1051 (1976).

138) C.A.Klein,
J.Appl.Phys., 39, 2029 (1968).

139) A.Rothwarf,
J.Appl.Phys., 44, 752 (1973).

140) J.R.Srour, O.L.Curtis,and K.Chiu,
IEEE Trans.Nuc.Sci. NS-21, 73 (1974).

141) S.K.Haywood, M.M.Heyns and R.F.DeKeersmaecker,
Presented at INFOS '85, Toulouse, April 16-18th. (1985).

142) Z.A.Weinberg,
Appl.Phys.Lett., 27, 437 (1975).

143) F.J.Grunthaner, B.F.Lewis, N.Zamini, J.Maserjian and A.Madhukar,
IEEE Trans.Nuc.Sci. NS-27, 1646 (1980).

144) M.Fischetti,
Phys.Rev.B, 31, 2099 (1985).

145) T.N.Theis, D.J.DiMaria, J.R.Kirtley and D.W.Dong,
Phys.Rev.Lett., 52, 1445 (1984).

146) M.M.Heyns,
Appl.Phys.Lett., 44, 202 (1985).

147) C.Svensson and I.Lundstrom,
J.Appl.Phys., 44, 4657 (1973).

148) M.Hirose, M.Morita and Y.Osaka,
Jap.J.Appl.Phys., 16, 561 (1977).

149) K.Kristiansen,
Vacuum, 27, 227 (1977).

150) H.J.DeWit and C.Crevecoeur,
Phys.Lett., A50, 365 (1974).

151) P.Solomon, N.Klein and M.Albert,
Thin Solid Films, 35, 321 (1976).

152) R.M.Hill and L.A.Dissado,
J.Phys.C, 16, 4447 (1983).

153) R.M.Hill and L.A.Dissado,
J.Phys.C, 16, 2145 (1983).

154) R.M.Hill, L.A.Dissado and R.J.Jackson,
J.Phys.C, 714, 3915 (1981).

156) L.A.Dissado, J.C.Fothergill, S.V.Wolfe and R.M.Hill,
IEEE Trans.Elec.Insulation EI-19, 227 (1984).

157) J.Mason,
Proc.IEEE, 128A, 193 (1981).

158) R.A.Metzler,
IEDM Tech.Dig., p233 (1979).

159) M.M.Heyns and R.F.DeKeersmaecker,
J.Appl.Phys., 58, 3936 (1985).

160) J.F.Verweij and D.R.Wolters,
Presented at INFOS'85 Toulouse, April 16-18th (1985).

161) A.Modelli and B.Ricco,
IEDM Tech.Dig., p148 (1984).

162) K.Hamano,
Jap.J.Appl.Phys., 13, 1085 (1974).

163) G.W.Brown,
IEEE Trans.Power App.Syst. PAS-101, 3513 (1982).

164) M.V.Fischetti,
J.Appl.Phys., 56, 575 (1984).

165) D.J.DiMaria, D.W.Dong, F.L.Pesavento, C.Lam and S.R.Brorson,
J.Appl.Phys., 55, 3000 (1984).

166) D.J.Robbins, D.J.DiMaria, C.Falcony and D.W.Dong,
J.Appl.Phys., 54, 4553 (1983).

167) A.K.Sinha and T.E.Smith,
J.Electrochem.Soc., 125, 743 (1978).

168) A.K.Sinha, H.J.Levinstein, L.P.Adda. E.N.Fuls and E.H.Povilonis,
Solid State Electron., 21, 531 (1978).

169) N.Shiono and C.Hashimoto,
IEEE Trans.Electron.Dev. ED-29, 361 (1982).

170) R.B.Fair and R.C.Sun,
IEEE Trans.Electron.Dev. ED-28, 83 (1981).

171) W.Meyer and R.B.Fair,
IEEE Trans.Electron.Dev. ED-30, 96 (1983).

172) P.V.Gray and D.M.Brown,
Appl.Phys.Lett., 8, 31 (1966).

173) Y.C.Cheng,
Prog. in Surface Sci., 8, 181 (1977).

174) Y.Tarui, K.Nagai and Y.Hayashi,
Electronics Comm. Jap., 53C, 146 (1970).

175) R.H.Walden,
J.Appl.Phys., 43, 1178 (1972).

176) A.I.Akinwande, C.P.Ho and J.D.Plummer,
Appl.Phys.Lett., 45 263 (1984).

177) F.Montillo and P.Balk,
J.Electrochem.Soc., 118, 1463 (1971).

178) G.Hu and W.C.Johnson,
Appl.Phys.Lett., 36, 590 (1980).

179) G.J.Hu and W.C.Johnson,
J.Appl.Phys., 54, 1441 (1983).

180) S.Yamasaki,
Jap.J.Appl.Phys., 10, 1555 (1971).

181) F.J.Grunthaner, P.J.Grunthaner, R.P.Vasquez, B.F.Lewis,
J.Maserjian and A.Madhukar,
J.Vac.Sci.Tech., 16, 1443 (1979).

182) F.J.Grunthaner, P.J.Grunthaner, R.P.Vasquez, B.F.Lewis,
J.Maserjian and A.Madhukar,
Phys.Rev.Lett., 43, 1683 (1979).

183) F.J.Grunthaner, P.J.Grunthaner, M.H.Hecht and D.Lawson,
Presented at INFOS '85, Toulouse, April 16-18th. (1985).

184) P.J.Burkhardt,
IEEE Trans.Electron.Dev. ED-13, 268 (1966).

185) F.M.Fowkes and D.W.Hess,
Appl.Phys.Lett., 22, 377 (1973).

186) M.Aslam and P.Balk,
"Insulating Films on Semiconductors", p103,
Eds. J.F.Verweij and D.R.Wolters, North Holland (1983).

187) K.Hubner,
J.Phys.C., 17, 6553 (1984).

188) Z.Weinberg, D.R.Young, D.J.Dimaria and G.W.Rubloff,
J.Appl.Phys., 50, 5757 (1979).

189) M.V.Fischetti, R.Gastaldi, F.Maggioni and A.Modelli,
J.Appl.Phys., 53, 3129 (1982).

190) Z.Weinberg and G.W.Rubloff,
Appl.Phys.Lett., 32, 184 (1978).

191) C.T.Sah, J.Y-C.Sun and J.J-T.Tzou,
J.Appl.Phys., 55, 1525 (1984).

192) P.J.Burkhardt,
J.Electrochem.Soc., 114, 196 (1967).

193) T.E.Burgess and E.M.Fowkes,
J.Electrochem.Soc., 113, 63C (1965).

194) S.R.Hofstein,
IEEE Trans.Electron.Dev. ED-14 749 (1967).

195) S.R.Hofstein,
IEEE Trans.Electron.Dev. ED-13 222 (1966).

196) G.L.Holmberg, A.B.Kuper and F.D.Miraldi,
J.Electrochem.Soc., 117, 677 (1970).

197) S.R.Hofstein,
Appl.Phys.Lett., 11, 95 (1967).

198) R.Bruckner,
J.Noncrystalline Solids, 5, 177 (1971).

199) F.A.Cotton and G.Wilkinson,
"Advanced Inorganic Chemistry", Wiley Interscience (1980).

200) M.W.Hillen,
Ph.D. thesis, Rijksuniversiteit Groningen, (1981).

201) V.Zekeriya and T.P.Ma,
IEEE NS-REC July 1984, Colorado Springs.

202) V.Zekeriya and T.P.Ma,
J.Appl.Phys., 56, 1017 (1984).

203) F.Gaensslen and J.Aitken,
IEEE Electron.Dev.Lett., EDL-1, 231 (1980).

204) P.Heremans, N.Saks, L.Van den Hove and H.Maes,
Presented at ESSDERC, Aachen, 9-12th. Sept. (1985).

Charge Build Up and Breakdown in Thin SiO_2 Gate Dielectrics

M. W. Hillen[+], R. F. De Keersmaecker[++], M. M. Heyns, S. K. Haywood[°], and I. S. Darakchiev[°°]

ESAT Laboratory
Katholieke Universiteit Leuven
Heverlee, Belgium

ABSTRACT

SiO_2 layers with low defect densities have been grown in a double-walled oxidation tube, for use as thin gate dielectrics in MOS IC's. Under all high-field stress conditions positively charged slow states are created at the Si-SiO_2 interface. These can be neutralized by applying positive fields at the gate of an MOS device. Negative charge can also be generated, especially during a positive field stress. The total amount of generated charge is much less for polysilicon gate capacitors than for Al-gate capacitors. The time-to-breakdown in a wearout experiment could be extended by periodic application of a positive gate voltage. This also caused neutralization of the slow states. However, it had no influence on the high-field breakdown distribution in a fast voltage ramp experiment. It is suggested that interface rather than bulk phenomena dominate trap generation and charge build-up during the high-field stresses which induce oxide breakdown.

1. INTRODUCTION

In integrated circuit technology improvements in speed, density, and yield can be obtained by reducing the dimensions of MOSFET's. The principle of scaling [1] offers the possibility of designing small-geometry transistors based on a knowledge of larger ones. In practice, for various reasons, voltages cannot be scaled down by the same factor as device dimensions. This leads to a number of effects on device characteristics related to high fields. As far as gate oxides are concerned, this means that they must withstand higher fields and that they are more susceptible to trapping of carriers from the substrate.

SiO_2 layers of 20 to 50 nm thickness were fabricated on Si for use as gate oxides in 0.5 to 3 μm channel-length MOS transistors. These oxides should have low charge densities, low trap densities (traps being any kind of electrically active microscopic defect), and low flaw (macroscopic defect) densities in order to have a high dielectric strength and a long time-to-breakdown under electrical stress.

It has been suggested that breakdown of SiO_2 layers on Si is preceded by injection and trapping of charge and by generation of trapping sites. For instance, generation of electron traps both in the bulk [2,3] and near the Si-SiO_2 interface [4] during high-field stressing has been inferred from *C-V* and ramp *I-V* measurements. These data led us to initiate the second part of the present study.

For the first time, accurate charge injection and sensing techniques have been applied before and after stress at low and high fields, to thin thermally grown oxides on n- and p-type silicon.

2. EXPERIMENTAL

<100> n- and p-type silicon wafers of doping levels 10^{15} and 10^{17} cm^{-3} were oxidized in dry O_2 at 900°C, which is an optimum temperature for the growth of oxides in the required thickness range. The oxidations were carried out in a double-walled furnace tube in order to meet the requirements of low charge, trap, and defect densities. After oxidation the wafers were

0018-9367/84/0600-0245$01.00 © 1984 IEEE

annealed for 10 minutes in N_2 at the oxidation temperature. On some of the wafers a 450 nm thick polycrystalline silicon layer was deposited in an LPCVD system. The polysilicon was then phosphorus doped from a liquid source and annealed at 975°C in N_2, to give a resistance of 20 Ω per square sheet.

MOS capacitors were formed by magnetron sputter deposition of 15 nm or 1 μm Al (depending on whether or not photo *I-V* measurements were to be applied) and then etching to give circular gates of various dimensions. The final anneal was 20 min at 435°C in forming gas (10% H_2 in N_2).

3. INFLUENCE OF PROCESSING ON DIELECTRIC STRENGTH

In the dielectric strength measurements the capacitors were subjected to a fast voltage ramp (ramp rate 100 V/s), while monitoring the current flowing into the oxide. When a preset current level (0.5 mA) was exceeded, the voltage was measured by a digital voltmeter connected to an automated data-acquisition system. When plotting the relative number of breakdowns as a function of the electric field, three groups were generally distinguished in the resulting histogram [5]: low-field breakdowns caused by gross defects which short the oxide layer, medium-field breakdowns caused by small defects, and high-field breakdowns tentatively ascribed to an intrinsic property.

The most important advantage of using a double-walled furnace tube instead of a conventional single-walled one is that fewer medium-field breakdowns occur (Fig. 1) [6].

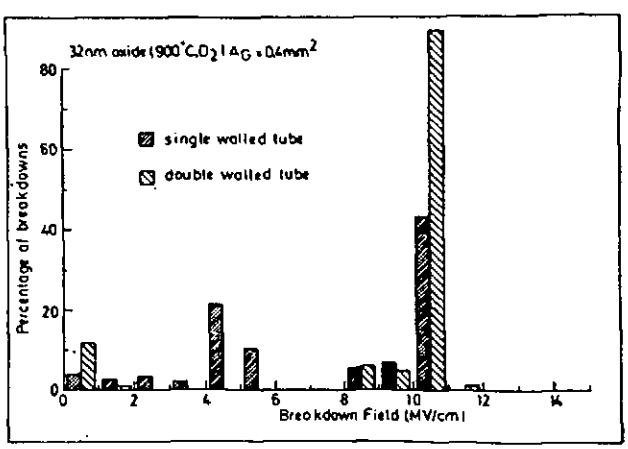


Fig. 1: Breakdown histograms of oxides grown in a single- and a double-walled oxidation tube.

The reason is, most probably, that metal ions (including alkali ions) which diffuse from the heating elements and the furnace ambient through the outer wall, are carried away by the gas flow in the outer tube and thus prevented from reaching the oxidation ambient.

In this experiment O_2 was used in the outer tube. When a small percentage of some chlorine containing compound was added to the gas flow in the outer tube we saw a narrower distribution of high-field breakdowns and a slight decrease of the number of low-field breakdowns (Fig. 2). These effects have been reported to occur for oxidations carried out in a chlorinated atmosphere at higher temperatures [7]. It is improbable that chlorine ions can diffuse through the inner furnace wall and we, therefore, suggest that chlorine in the outer tube prevents some species from diffusing into the oxidation ambient. This species is thought to be responsible for at least part of the high-field breakdowns in the experiment without chlorine, which implies that high-field breakdowns are not caused solely by an intrinsic process.

The use of a higher flow rate in the oxidation tube increased the maximum breakdown field. This is shown in Fig. 3, where the cumulative percentage of breakdowns is plotted as a function of field for a capacitor with a very small area (for which the chance of finding a low-field breakdown is correspondingly small). It is believed that the higher gas flow may carry off unwanted particulates more effectively, preventing the formation of tiny defects in the oxide which trigger the final breakdown.

The use of phosphorus-doped polysilicon electrodes led to a higher maximum breakdown field (Fig. 4).

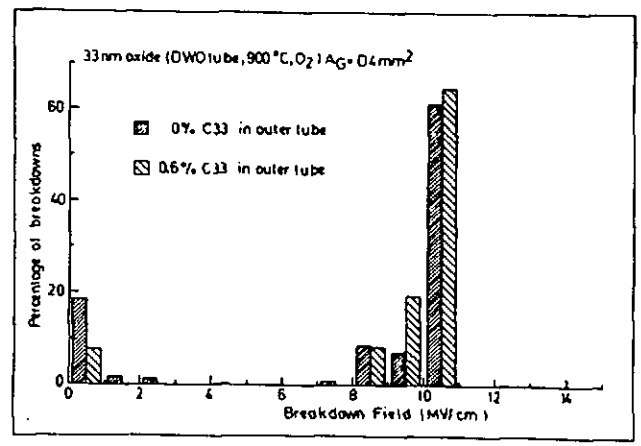


Fig. 2: Breakdown histograms of oxides grown in a double-walled tube. Gas flow in outer tube is O_2 or C33/O_2 mixture (C33 : 1,1,1-trichloroethane).

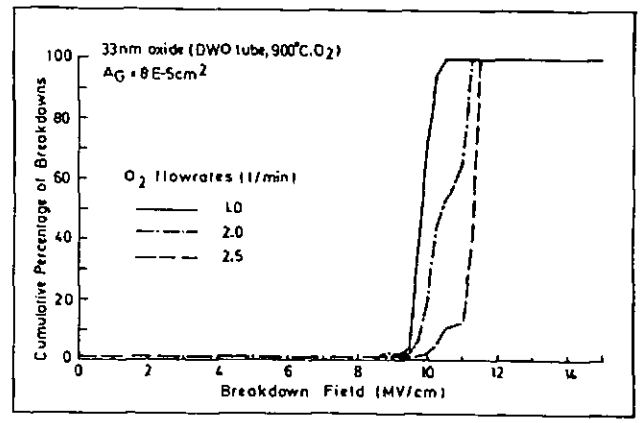


Fig. 3: Cumulative breakdowns as a function of field for different O_2-flow rates in the tube.

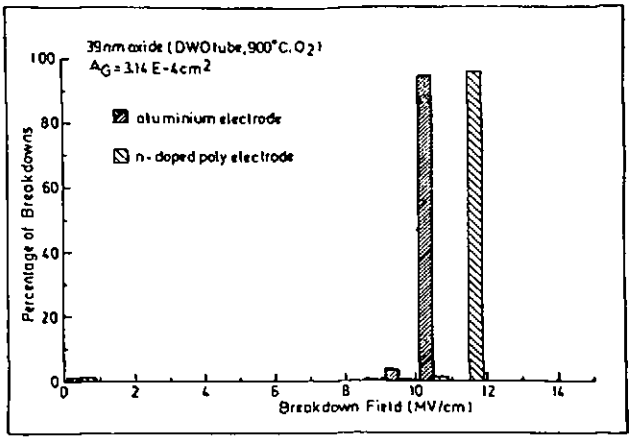


Fig. 4: Breakdown histograms of capacitors with aluminum and polysilicon electrodes.

4. CHARGE TRAPPING IN ALUMINUM AND POLY-SILICON CAPACITORS

A first impression of the charge build-up during stressing with a rapidly increasing field is obtained from ramp *I-V* measurements. In this technique the current response to a voltage ramp of the order of 1 V/s is measured with a logarithmic current-to-voltage convertor. The current is determined by Fowler-Nordheim injection of electrons from the negatively biased electrode, which in turn is influenced by the sign, amount, and location of trapped charge [3]. Results from more accurate and sensitive charge measurements are reported in the next section.

Charge trapping during high-field stressing was studied using the ramp *I-V* method. We used capacitors with very small electrode areas and oxides grown in the double-walled tube in order to obtain almost exclusively high-field breakdowns. Typical *I-V* curves on capacitors with aluminum and polysilicon electrodes are shown in Fig. S. Curve 1 represents the well known Fowler-Nordheim injection of electrons. When the voltage ramp was stopped before breakdown and the measurement repeated immediately, curve 2 was obtained. The threshold voltage for electron injection is reduced by positive charge near the $Si-SiO_2$ interface and at higher current levels a shift to higher gate voltages is seen, caused by trapped electrons.

When the voltage ramp was stopped as in the previous experiment and the measurement repeated after a stress at positive voltage (S MV/cm for S minutes) curve 3 was recorded: most of the positive charge has disappeared and only negative trapped charge remains. The amount of trapped charge is dependent on the exact stress conditions but is always much less for capacitors with polysilicon electrodes. This is in agreement with earlier conclusions from avalanche injection experiments [8].

It should be noted that quantitative information about the charge trapped in the oxide during stress cannot be obtained from ramp *I-V* measurements, because the high field and the injected current tend to disturb the charge distribution already present in the oxide. In the next section photo *I-V* measurements will be used to obtain more reliable information on oxide charge distributions.

5. PHOTO I-V MEASUREMENTS

To gain more insight into the type and location of the charge build-up during constant stress experiments, internal photoemission measurements as a function of voltage (photo *I-V*) [9] were carried out for the MOS capacitors with a transparent aluminum gate. A 150 W Xenon arc lamp and a 25 cm monochromator were used and the photocurrents were measured using a high-speed electrometer and an automatic data acquisition system.

Figure 6 shows photo *I-V* curves from a 39 nm oxide on p-type silicon before and after a constant current stress at negative gate voltage. The flat-band voltage shift ΔV_{fb} after this stress was -12V, but no shift can be seen in the photo *I-V* curves, indicating that the positive charge is located at (within 1 nm of) the $Si-SiO_2$ interface. This positive charge can be attributed to (slow) states located at the $Si-SiO_2$ interface, which are created in a positive charge state but can be neutralized easily by applying a moderate positive voltage stress. Similar states were reported to be generated in electron avalanche injection experiments [10]. Such a positive voltage also occurs during the positive part of a *C-V* recording [6] or during the retrace in a ramp *I-V* measurement, when a triangular ramp is used. Slow states were detected upon the application of both positive and negative stresses to oxides on p-type as well as n-type silicon. The dispersion at high positive voltage for the curves in Fig. 6 before and after stress, indicates the presence of some negative charge close to (within 3 nm of) the $Si-SiO_2$ interface. Photo *I-V* measurements after avalanche injection [11] of electrons from the substrate at an oxide field of about 4 MV/cm showed that extra electron traps (compared to an unstressed sample) are created close to the $Si-SiO_2$ interface. Similar results were found after negative voltage stresses for oxides on n-type silicon.

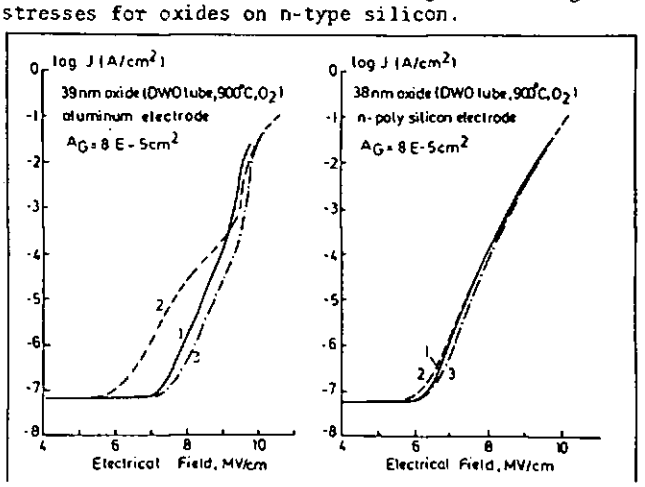


Fig. 5: Ramp I-V curves for (a) oxide on n-type Si with Al electrode and (b) oxide on n-type Si with poly-Si electrode. Explanation in text.

Photo *I-V* curves before and after positive current stresses on p-type samples (d_{ox}=39 nm) are shown in Fig. 7. The curve recorded after application of the stress shows negative charge near the $Si-SiO_2$ interface and positive charge near the $Al-SiO_2$ interface, possibly due to positively charged Al-related centers.

From an appreciable negative ΔV_{fb} after this stress it was concluded that again slow states are generated at the $Si-SiO_2$ interface. The photo *I-V* curve after avalanche injection of electrons shows that no electron traps are created close to the Si-SiO interface.

Similar results were obtained on n-type samples using a positive current stress.

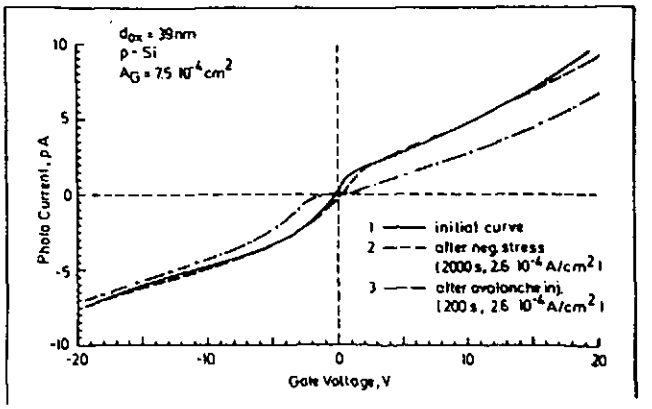


Fig. 6: Photo I-V curves of p-type MOS capacitor with Al electrode before and after negative current stress.

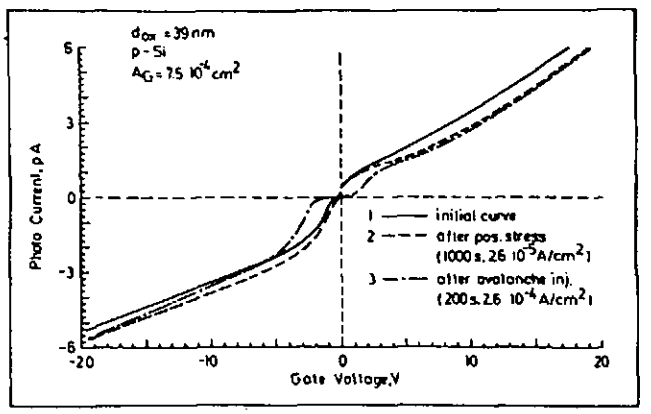


Fig. 7: Photo I-V curves of p-type MOS capacitor with Al electrode before and after positive current stress.

6. STRESS EXPERIMENTS

The flat-band voltage shift as a function of total injected charge is depicted in Fig. 8 for a 35 nm oxide on n-type silicon (Al gate) during different constant current strasses at positive gate voltages (called positive current stresses). The neutralization of the positively charged oxide states by the applied positive voltage (36 to 40 V) is reflected in the increasing value of ΔV_{fb} and is more efficient at lower injection current densities.

For p-type samples we found an increasing negative ΔV_{fb} for stresses of both polarities [6]. The fact that positively charged states are not neutralized efficiently for p-type samples in inversion (positive gate bias) is related to the low generation rate of minority carriers.

From Fig. 8 it follows that the time-to-breakdown t_{BD} in a wear-out experiment (prolonged application of constant current or voltage stress) is not a unique function of the amount of injected charge, as was suggested previously [12]. We have found that t_{BD} can be increased considerably by interspersing periods of positive bias with the constant current injection. During these periods of positive bias the slow states are also largely neutralized.

Fast voltage ramp dielectric strength measurements are not affected by the presence and charge state of these states. We found similar breakdown voltages for unstressed samples and for stressed samples, regardless of whether the slow states have been previously charged positively or neutralized.

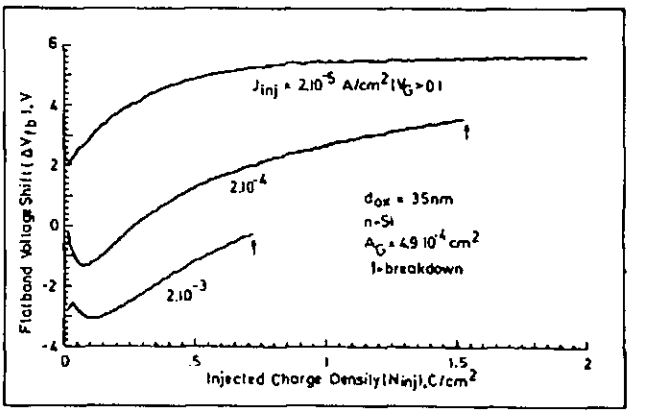


Fig. 8: Flat-band voltage shift as a function of injected charge for positive current stress.

7. DISCUSSION AND CONCLUSIONS

SiO_2 layers on Si with extremely low defect densities have been grown in a double-walled oxidation tube. The use of a chlorinated gas in the outer tube, the increase of O_2 flow in the inner tube, and the use of polysilicon electrodes all improve the maximum breakdown fields.

The charge distributions found in the oxides after high-field stressing are depicted schematically in Fig. 9. To a first approximation the distribution obtained is only a function of the applied voltage polarity. An important observation is that no bulk charge exists after high-field stressing. The reason is that the large electric field in the oxide bulk detraps all charge immediately [13]. Electrons can eventually become trapped near the oxide interfaces, especially near the Si-SiO_2 interface in positively stressed samples (where they constitute negative interface charge). In general less charge is trapped in polysilicon capacitors than in Al-gate capacitors.

Slow states are created at the Si-SiO_2 interface under all stress conditions. They can be neutralized by applying positive fields. The presence and charge state of the slow states has no influence on the final breakdown voltage in a dielectric strength measurement. However, the intermittent application of a positive voltage during a wearout experiment caused the neutralization of the slow states and increased the time-to-breakdown.

Under negative high-field stress defects are created near the Si-SiO_2 interface, which could indicate the onset of the final breakdown process. We expected the existence of similar defects near the electrode-SiO_2 interface in positively stressed samples, but in Al-SiO_2-Si capacitors such defects could not be detected because of the dominant electron trapping by positive charge created under the Al-gate. Experiments with polysilicon gate structures are on the way.

8. ACKNOWLEDGMENT

We thank M. Reybrouck for carefully preparing the samples.

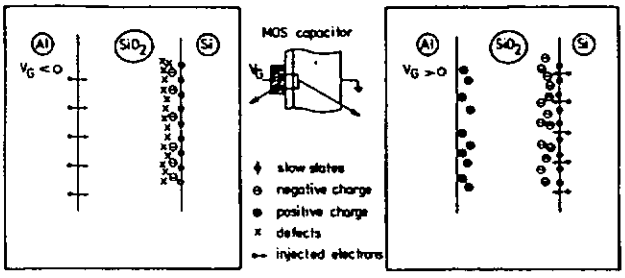


Fig. 9: Schematic of charge distribution in Al-SiO_2-Si capacitors after negative and positive current stress.

REFERENCES

[1] R. H. Dennard, F. H. Gaenslen, H.-N. Yu, V. L. Rideout, E. Bassous and A. R. Le Blanc, IEEE J. Solid-State Circuits SC-9, 256 (1974).

[2] C. S. Jenq, T. R. Ranganath, C. H. Huang, H. S. Jones and T. T. L. Chang, IEDM 81, Technical Digest, p. 388 (1981).

[3] P. Solomon, J. Appl. Phys. 48, 3843 (1977).

[4] E. Harari, J. Appl. Phys. 49, 247B (1978).

[5] N. J. Chou and J. M. Eldridge, J. Electrochem. Soc. 117, 1287 (1970).

[6] M. W. Hillen, R. F. De Keersmaecker, M. M. Heyns, S. K. Haywood and I. S. Darakchiev, in "Insulating Films on Semiconductors - INFOS 83", (J.F. Verwey and D. R. Wolters, eds.), North-Holland, Amsterdam (1983), p. 274.

[7] C. M. Osburn, J. Electrochem. Soc. 121, 809 (1974).

[8] D. R. Young, E. A. Irene, D. J. DiMaria, R. F. De Keersmaecker and H. Z. Massoud, J. Appl. Phys. SD 6366 (1979).

[9] D. J. DiMaria, J. Appl. Phys. 47, 4073 (1976).

[10] M. Heyns and R. F. De Keersmaecker, Paper presented at the joint session of the 11th European Solid-St. Device Research Conf. and the 6th Symp. on Solid-St. Device Technology, Sept. 14-17, 1981 (unpublished).

[11] E. H. Nicollian, A. Goetzberger and C. N. Berglund, Appl. Phys. Lett. 15, 174 (1969).

[12] O. R. Wolters, in "Insulating Films on Semiconductors - INFOS 81" (M. Schulz and G. Pensl, eds.), Springer, Berlin (1981) p. 18D. p. 111.

[13] M. M. Heyns, R. F. De Keersmaecker and M. W. Hillen, Appl. Phys. Lett. 44, 202 (1984).

+ *Present address: Philips Elcoma, Nijmegen The Netherlands*

++ *Research Associate of the National Fund for Scientific Research (Belgium)*

° *Permanent address: Microelectronics Centre, Middlesex Polytechnic, London, England*

°° *Under grant of State Committee for Science and Technical Progress, Sofia, Bulgaria.*

This is a combined version of two papers presented at the First International Conference on Conduction and Breakdown in Solid Dielectrics, Toulouse, July 4-8, 1983.

Hole trapping and interface state generation during bias-temperature stress of SiO_2 layers

S. K. Haywood and R. F. De Keersmaecker
E.S.A.T. Laboratory, Kathalieke Universiteit Leuven, Kardinaal Mercierlaan 94, 3030 Heverlee, Belgium

(Received 7 May 1985; accepted for publication 29 May 1985)

Avalanche hole injection measurements on unstressed metal-oxide-semiconductor capacitors and those having undergone bias-temperature stress (BTS) are compared. The results show that the negative midgap voltage shift (ΔV_{mg}) occurring during negative BTS is due to filling of intrinsic hole traps. Conversely, positive BTS removes any previously trapped holes from the oxide. No evidence was found for hole trap generation at either stress polarity. Interface state generation across the band gap accompanies hole trapping during negative BTS and a characteristic peak is generated in the interface trap distribution at ~ 0.2 eV above midgap on neutralization or detrapping of these holes during positive BTS.

A shift in threshold voltage can seriously affect the reliability of a metal-oxide-semiconductor field-effect transistor (MOSFET). In SiO_2 layers free of mobile ions, one common form such a shift may take is the negative bias instability. This is a negative shift of the threshold voltage, V_T, following prolonged operation under negative gate bias. Under real operating conditions this is a long-term instability. The aging process can be accelerated however, by applying the bias at an elevated temperature, the so-called bias-temperature stress (or BTS) measurement. The phenomenon is also more severe in aluminum than polycrystalline silicon gate devices. For this reason the measurements presented here will be on aluminum gate capacitors.

Over the past 20 years the negative bias instability has been extensively studied, both from the point of view of kinetics[1–11] and of processing dependence.[1,2,4,9,10] Various models have been proposed for the positive charge generation; these include hole trapping[3] or electron emission from traps in the oxide,[5,6] the breaking of weak bonds near the Si-SiO_2 interface,[1] and the formation and migration of oxygen vacancies.[12] In this work an attempt was made to shed more light on the means of positive charge generation by elucidating the role of hole traps. Interface state generation accompanying BTS and its relationship to hole trapping was also examined.

Capacitors were fabricated on n-type ⟨100⟩ silicon wafers of (1–2) $\times 10^{17}$ cm^{-3} doping. This heavy doping is necessary to allow laterally homogeneous avalanche injection.[13] Oxidation was carried out in a double-walled furnace at 900 °C in dry O_2.[14] SiO_2 layers of 25–55 nm were grown and the wafers then received a nitrogen anneal, also at 900 °C for 15 min. An aluminum layer, usually 1 μm thick, was deposited by dc magnetron sputtering and electrodes were defined by standard wet lithography. Finally the samples received a post-metallization anneal in forming gas at 435 °C for 20 min.

BTS measurements were conducted at 100 °C $< T <$ 305 °C and fields of 1–4 MV cm^{-1} for times of up to 20 h. Samples were heated and cooled under bias and all C-V curves were measured at room temperature. A positive BTS of 2 MV cm^{-1} at 250 °C for 15 min (2 MV cm^{-1}/250 °C/15 min) showed that the mobile ion concentration was below 10^{10} cm^{-2}.

Avalanche injection of holes from the silicon into the SiO_2 using a sawtooth waveform[15] was performed on unstressed capacitors and on devices having already undergone BTS. The injection current, measured with an electrometer (Keithley 616), was kept constant and the midgap voltage shift ΔV_{mg} was measured at intervals using a capacitance meter (Boonton 72BD) to see if any hole traps had been filled or generated during the BTS.

The interface state density distribution over the band gap was calculated from the difference between the quasi-static and high-frequency-capacitance-voltage (C-V) curves in depletion and from the quasi-static and ideal low-frequency curves over the rest of the band gap.

Throughout the present work, ΔV_{mg} has been taken as a measure of the change in positive oxide charge alone, with no contribution from interface states. If above midgap the interface states are all acceptors and below midgap they are all donors (or indeed vice versa providing there are equal numbers of acceptor and donor states) the midgap voltage is a neutral point in the interface state distribution and any shift in this voltage, ΔV_{mg}, is representative only of a change in the amount of oxide charge. This has been assumed to be the case by several authors for interface states generated by different types of stress: avalanche hole injection,[16] negative BTS,[7,11] high-field stress,[11] and irradiation.[11,17] As will be shown here this assumption appears to be valid for BTS induced interface states.

Figure 1 shows the shift of the high-frequency C-V curve following application of a negative BTS of -4 MV cm^{-1}/250 °C/1h to a 28 nm oxide, curve (b). On reversing the stress ($+4$ MV cm^{-1}/250 °C/1h) the C-V curve moves back to more positive voltages, curve (c). It will be shown later from avalanche hole injection measurements that all the positive charge generated by negative BTS is removed by the subsequent positive BTS. There is considerable interface state generation following the positive BTS, however, causing distortion of the C-V curve. Nevertheless, curve (c) in Fig. 1 crosses the initial C-V curve [curve (a)] at V_{mg}. Hence the net charge in the interface states is zero at midgap and ΔV_{mg} can be used as a measure of the change in oxide charge alone.

An indication that the negative bias instability might be due to hole trapping came from an observed correlation between the magnitude of ΔV_{mg} seen on a sample subjected

 0003-6951/85/160381-03$01.00

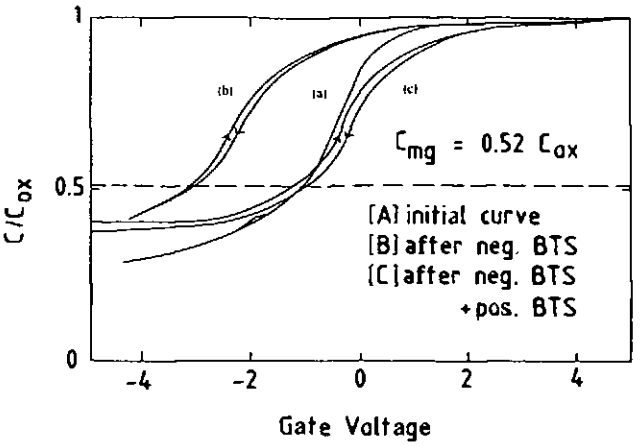


FIG. 1. High-frequency C-V curves taken at 1 MHz following BTS on a 28 nm oxide of gate area 1.08×10^{-3} cm^2. Curve (a) before stress; curve (b) after -4 MV cm^{-1}/250 °C/h; curve (c) after -4 MV cm^{-1}/250 °C/h plus $+4$ MV cm^{-1}/250 °C/h.

to negative BTS and ΔV_{mg} seen during avalanche hole injection.

To examine this idea further, avalanche hole injection was carried out on a sample having undergone negative BTS and the evolution of ΔV_{mg} compared with that during hole injection on a virgin sample. If ΔV_{mg} during negative BTS were entirely due to hole trapping, a subsequent hole injection measurement would show the same saturation value of the *absolute* midgap voltage V_{mg} as is observed on a virgin sample. On the other hand ΔV_{mg}, the midgap voltage *shift*, occurring during hole injection would differ in the two cases by the magnitude of ΔV_{mg} already found during negative BTS. This is, in fact, what was observed and the results are illustrated in Fig. 2. Figure 2(a) shows ΔV_{mg} during the hole injection for a virgin sample (—) and for one having undergone a stress of -3 MV cm^{-1}/250 °C/5.5 h (----). The difference in the saturation value of the two curves tends towards the value of ΔV_{mg} already present at the moment

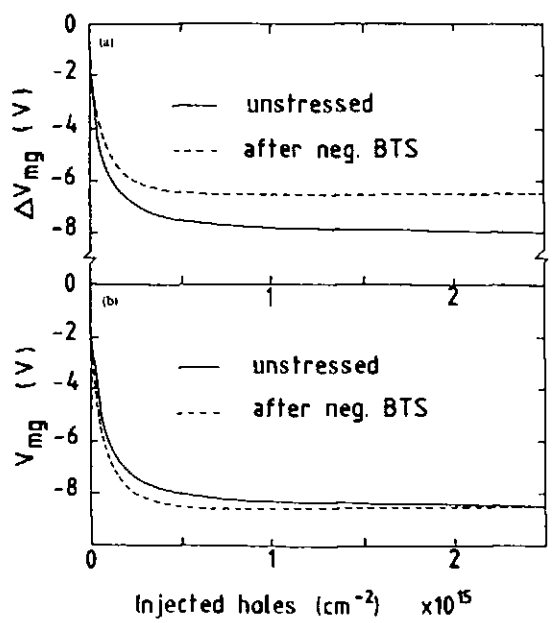


FIG. 2. Evolution of (a) ΔV_{mg} and (b) V_{mg} during avalanche injection of holes into a 28 nm oxide at a current density of 9.3×10^{-8} A cm^{-2}. (solid lines) unstressed sample; (dashed lines) after BTS of -3 MV cm^{-1}/250 °C/5.5 h.

hole injection commenced; in this case approximately -2 V. Figure 2(b) shows the absolute value of the midgap voltage in the two cases. Here both curves tend towards the same saturation value.

As can be seen from Fig. 1, positive BTS after negative BTS returns V_{mg} to very close to its initial value. Subsequent avalanche hole injection resulted in a ΔV_{mg} versus injected charge curve very similar to that obtained on a virgin sample. Hence the positive stress removed all the holes from the oxide; no new traps were created during BTS at either polarity and there was also no change in capture cross section or trap density as a result of hole trapping. It would appear therefore that hole trapping during negative BTS is a completely reversible process. It is, however, accompanied by significant interface degradation which will now be discussed in more detail.

Interface state generation was seen across the whole band gap following negative BTS. As has been observed by other workers[7] the change in midgap interface state density, ΔD_{it} (midgap), during negative BTS was proportional to ΔV_{mg} i.e., to the number of trapped holes (Fig. 3(a)]. We found the ratio of interface states to trapped holes to be greater than the value of 1 reported in Ref. 7. Some of the positive charge is very easily detrapped (e.g., by the taking of a C-V curve) and this apparently high ratio is probably due to an underestimation of ΔV_{mg}.

Lai reports a linear relationship between holes trapped and ΔD_{it} (midgap) during avalanche hole injection.[16] This was confirmed by our measurements. ΔD_{it} (midgap) saturated with injected charge but was a linear function of ΔV_{mg} [see Fig. 3(b)]. In this case the ratio of interface states at midgap to holes trapped was measured as ~0.5. The charge may well be trapped further into the oxide during avalanche injection where there is a substantial hole current flowing

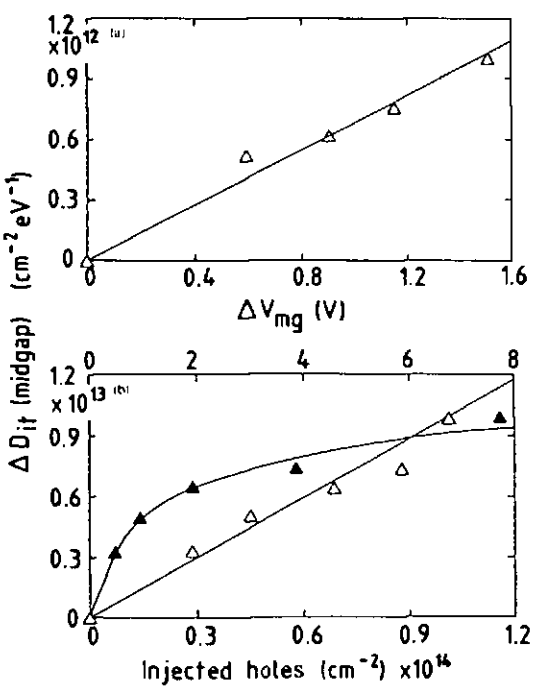


FIG. 3. (a) Growth of ΔD_{it} (midgap) as a function of ΔV_{mg} during negative BTS on a 39 nm oxide. (b) Growth of ΔD_{it} (midgap) during avalanche hole injection into a 28 nm oxide at a current density of 4.65×10^{-8} A cm^{-2}. (△) vs ΔV_{mg} and (▲) vs injected charge.

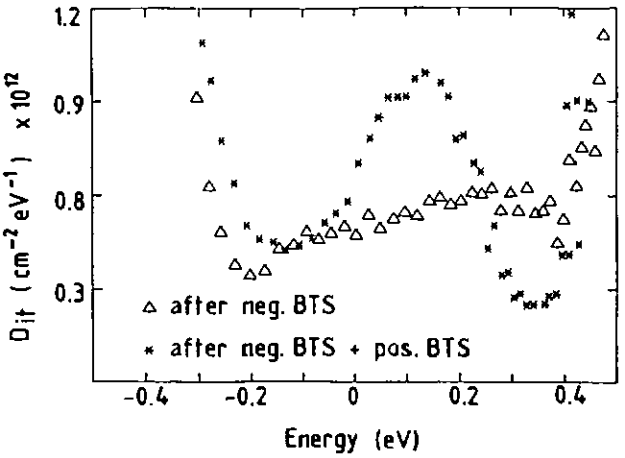


FIG. 4. Interface state generation over the silicon band gap during BTS on a 39 nm oxide (0 = midgap). (△) After -2 MV cm^{-1}/250 °C/15 min; (*) after -2 MV cm^{-1}/250 °C/15 min plus $+2$ MV cm^{-1}/250 °C/15 min.

than is the case for negative BTS where no measurable current flow was observed once the set temperature had been reached. Holes trapped deeper into the oxide will be less easily removed thus giving a larger measured value of ΔV_{mg} for a given value of ΔD_{it} (midgap).

Removal of holes from the oxide also resulted in significant interface degradation. If negative BTS was followed by positive BTS, then a peak was generated in the interface state distribution at between 0.15 and 0.25 eV above midgap (Fig. 4). The precise position varied somewhat with processing and to a lesser extent ($\pm$ 0.02 eV) with the parameters chosen for the C-V analysis. Lai has shown using avalanche hole injection followed by photoinjection of electrons that generation of such a peak in the interface state density distribution is a two-stage process accompanying the capture of electrons by trapped holes.[16] Other forms of stress which supply both electrons and holes to the oxide, such as high-field injection and irradiation, also result in generation of this peak.[11] It appears that a similar two-stage process of interface state generation occurs during hole trapping and neutralization or detrapping in the negative BTS/positive BTS sequence.

The peak in the interface state distribution was confirmed to be genuine and not due to lateral nonuniformities since it appeared both in the distribution measured by the C-V technique on stressed capacitors and in that measured by charge pumping[18,19] on similarly stressed transistors on the same wafer. The latter technique directly measures a current from the charging and discharging of the interface states and is not sensitive to fluctuations in the surface potential.

The precise mechanism by which holes enter the oxide under BTS conditions is not known. The most likely mechanism seems to be direct hopping of holes, already present at the Si-SiO_2 interface under the negative bias, into an oxide hole trap. Such a mechanism would not cause a measurable current in the outer leads if the hole traps are present close to the Si-SiO_2 interface. A hopping process is also consistent with the observed field and temperature dependence of the hole trapping during BTS. Discharging under positive bias may occur by either detrapping of the holes or direct tunneling of electrons to the trapped holes which act as Coulombic attractive traps. Since the discharging is also temperature dependent, the former mechanism seems most likely.

Since no hole trap generation was seen during these experiments and all the observed degradation was preceded by the filling of intrinsic hole traps, the number of such sites in an oxide is obviously an important property. There is evidence that filling of hole traps is also responsible for positive charge generation during other forms of stress.[20] Hole traps must therefore be taken into account when considering oxide quality along with the more frequently considered electron traps.

We would like to thank M. M. Heyns for many valuable discussions and a critical reading of the manuscript, G. Groeseneken and P. Heremans for assistance with the charge pumping measurements, and M. Reybrouck for sample preparation.

[1]B. E. Deal, M. Sklar, A. S. Grove, and E. H. Snow, J. Electrochem. Soc. **114**, 266 (1967).
[2]D. J. Breed and R. P. Kramer, Solid State Electron. **19**, 897 (1976).
[3]S. R. Hofstein, Solid State Electron. **10**, 657 (1967).
[4]A. Goetzberger, A. D. Lopez, and R. J. Strain, J. Electrochem. Soc. **120**, 90 (1973).
[5]D. J. Breed, Solid State Electron. **17**, 1229 (1974).
[6]D. J. Breed, Appl. Phys. Lett. **26**, 116 (1975).
[7]K. O. Jeppson and C. M. Svensson, J. Appl. Phys. **48**, 2004 (1977).
[8]A. K. Sinha and T. E. Smith, J. Electrochem. Soc. **125**, 743 (1978).
[9]A. K. Sinha, H. J. Levinstein, L. P. Adda, E. N. Fuls, and E. H. Povilonis, Solid State Electron. **21**, 531 (1978).
[10]N. Shiono, O. Nakajima, and C. Hashimoto, J. Electrochem. Soc. **130**, 138 (1983).
[11]G. Haller, M. Knoll, D. Bräunig, F. Wulf, and W. R. Fahrner, J. Appl. Phys. **56**, 1844 (1984).
[12]J. Bennett and L. M. Roth, J. Phys. Chem. Solids **32**, 1251 (1971).
[13]A. Rush and C. Bulucea, IEEE Trans. Electron Devices **ED-26**, 201 (1979).
[14]M. W. Hillen, R. F. De Keersmaecker, M. M. Heyns, S. K. Haywood, and I. S. Darakchiev, IEEE Trans. Electron. Insulation **EI-19**, 245 (1984).
[15]J. M. Aitken and D. R. Young, IEEE Trans. Nucl. Sci. **NS-24**, 2128 (1977).
[16]S. K. Lai, J. Appl. Phys. **54**, 2540 (1983).
[17]G. A. Scoggan and T. P. Ma, J. Appl. Phys. **48**, 294 (1977).
[18]J. S. Brugler and P. G. A. Jespers, IEEE Trans. Electron Devices **ED-16**, 297 (1969).
[19]G. Groeseneken, H. E. Maes, N. Beltran, and R. F. De Keersmaecker, IEEE Trans. Electron Devices **ED-31**, 42 (1984).
[20]M. M. Heyns (unpublished).