Integrated Processing of MOS Gate Dielectric Structures

G. W. Rubloff, M. Offenberg, and M. Liehr

Abstract—Integrated processing of MOS gate dielectric structures has been carried out in an ultraclean, multichamber processing system as a prototype for cluster (multichamber) processing in order to understand the potential and limitations of cluster processing for fabricating and processing for this application. With the enhanced surface cleanliness and contamination control enabled in this process environment, etching of the clean Si surface can occur when very low concentrations of oxygen species impinge on the surface at elevated temperatures, an intrinsic consequence of Si-O chemistry. Such etching leads to statistical roughening of the Si surface topography and thereby degrades the electrical properties of subsequently fabricated MOS structures. Identification of the etching reaction leads directly to prescriptions for cluster process integration which prevent etching: (1) ramping up to thermal oxidation temperature in an ambient with sufficient oxygen; and/or (2) in-situ formation of a passivating oxide surface layer as part of the pre-oxidation cleaning step.

I. INTRODUCTION

The FET gate dielectric structure is one of the most critical elements in advanced CMOS technology because it requires thermal oxide layers well under 10 nm, highly perfect interface and film electrical properties, high reliability, and acceptable yield and very high levels of integration. Manufacturability of the FET gate faces profound obstacles, particularly in contamination and defect density control. Although particulate contamination normally dominates such concerns, reactive impurities (primarily molecular species) may also be a crucial factor in determining the quality and yield of FET gate dielectric structures, especially with thiner structures and new process and tool elements. Integrated processing, also known as cluster or multichamber processing, has become an important trend in processing tools for manufacturing, driven in large part by their potential advantage in defect density control. The variety of commercially available integrated vacuum processing tools has now embraced processes normally carried out in hot wall batch reactors, either at low pressure (e.g., LPCVD polySi, nitride, oxide) or at atmospheric pressure (e.g., thermal oxidation). This opens the door to integrated processing of FET gate dielectric structures, which requires pre-oxidation cleaning, thermal oxidation, annealing, and polySi deposition.

Process, tool, and process integration issues abound for this cluster sequences, such as wet vs. vapor cleaning, or single-wafer rapid thermal vs. hot wall batch processing. Even before these can be examined, however, it seems prudent to assess the potential and limitations of the conventional process combination when integrated into a cluster tool and exercised under ultraclean conditions. The conclusion from the present exercise is that new chemical reaction pathways may become accessible in advanced cluster processing, and these must be identified and taken properly into account in defining a manufacturable process integration sequence for the cluster.

We have employed an advanced experimental research system, [1], [2] to address the underlying issues in integrated processing of FET gate dielectric structures. This tool includes a variety of process reactors as well as in situ analysis chambers, all ultrahigh-vacuum (UHV) so that reactive impurities can be highly controlled to allow a thorough assessment of their role in affecting advanced chemical processes, including integrated surface cleaning, thermal oxidation, annealing, and polySi deposition. The results demonstrate that the etching reaction—in which low oxygen concentrations etch an exposed Si surface and cause roughness—must be prevented by a suitable choice of ramp-up procedure before oxidation (in sufficient oxygen) and/or by formation of an oxide passivation layer before ramp-up [3]–[5].

Following a brief summary of our previous results, in this paper we present direct chemical evidence that the etching reaction is responsible for the systematic degradation of electrical breakdown (for Al gate MOS capacitors), show that the mechanism is also operative when pre-oxidation cleaning, thermal oxidation, and polySi gate deposition are all integrated into a cluster process, and provide prescriptions—with their rationale—for cluster process integration in an ultraclean tool set.

II. EXPERIMENT

In the present studies, a hot wall UHV-based quartz reactor was employed for thermal oxidation and annealing at atmospheric pressure as well as for in situ deposition of polySi from SiH4 at low pressure. The reactor is a vertical batch reactor for a small cassette of four 3.25 in diameter wafers. It is pumped by a turbomolecular pump system to reach < 1 × 10^-7 torr base pressure after bakeout, consisting mainly of H2, CO, and CO2; the total base pressure with the furnace at oxidation temperature is ~ 1 × 10^-6 torr. Purified Ar at 1 atm. was used in some cases during ramp-up to oxidation temperature as well as in annealing processes prior to oxidation. By introducing small concentrations (~ 0.1 - 1000 ppm) of O2 during preannealing in 1 atm. purified Ar, it was possible to ascertain the role of reactive impurities within the integrated gate dielectric process sequence. The hot wall reactor is connected to UHV transfer chambers (< 1 × 10^-9 torr) which permit wafer transfer without air exposure to other tools, including an inert-atmosphere glove box for pre-oxidation surface cleaning and various surface analysis instruments.

0894-6507/94$04.00 © 1994 IEEE
Si wafers (3.25 in dia., p-Si(100), B-doped, 2 Ω-cm) were pre-cleaned by standard RCA techniques, then introduced through a load-lock into a N₂-purged glove box for final pre-oxidation cleaning treatments; after the final clean, they were transferred into a vacuum (UHV) load-lock (~1 x 10⁻⁶ torr base pressure) and pumped down by sorption and cryopumps at a rate slow enough to prevent particle formation associated with condensation. The wafers were then moved in a TiN-coated cassette through the wafer transport system to the hot wall reactor, where they were transferred into a quartz cassette used during reaction.

A variety of final precleaning treatments were utilized in concert with the subsequent integrated thermal oxidation and annealing cycles. The following included processes which leave an oxide-covered Si surface are:

- Standard RCA clean,
- UV-ozone treatment to remove surface carbon (10' in O₂ with Hg radiation),
- Immersion in hot H₂O₂ (60°C, 10', in 1:5 H₂O₂: H₂O solution),

and also processes which leave the surface essentially oxide-free:

- Wet HF dip (10' in 1:10 HF solution),
- Vapor HF exposure (5' over vapors from 50% HF solution),
- Low temperature UHV/CVD growth of epitaxial Si (~ 1 mtorr SiH₄ at 550°C in ultrahigh-vacuum-based reactor),
- High temperature annealing in 1 atm Ar (> 900°C).

Following the final surface cleaning and transfer into the hot wall reactor, the wafers were ramped up to oxidation temperature in either Ar or O₂ at 1 atm. In some cases a preanneal in Ar with intentional, low concentrations of O₂ impurities was used. Then oxidation was carried out at ~ 850°C for 1 h to achieve ~ 12 nm SiO₂. MOS capacitor test sites were then fabricated either by in-situ polySi deposition (SiH₄ reaction under UHV/CVD conditions) and subsequent lithography and processing or by removing the wafers and carrying out Al dot evaporation through a shadow mask in another tool; in some cases, Al evaporation through a shadow mask was done in the integrated tool, with similar electrical results.

III. RESULTS AND DISCUSSION

Important conclusions from these experiments using Al-gate MOS capacitors have been published previously [3]–[15]. The most important of these can be summarized briefly as follows.

- High quality Al-gate MOS capacitors (e.g., 12 MV/cm breakdown field) are reproducibly obtained using ultraclean, integrated processing. This establishes credibility for the research tooling described above in prototyping the integrated processing of FET gate dielectric structures.
- For ramp-up to oxidation temperature in ultrapure Ar, final precleans which leave oxide on the surface result in high quality MOS capacitors (high breakdown fields), whereas poor breakdown characteristics are obtained when the surface is left oxide-free.
- Good breakdown behavior can be achieved using oxide-free surfaces by ramping-up in O₂ rather than in Ar.
- Significant evidence for Si surface roughening is found under conditions where low field breakdown is symptomatic. Structural evidence for roughness has been obtained from cross-sectional transmission electron microscopy and from scanning tunneling microscopy, while electrical evidence has been found in current-field characteristics, which show lower voltage turn-on of Fowler–Nordheim tunneling consistent with field-enhancement at asperities.

These results have been interpreted in terms of the etching of a Si surface which occurs at elevated temperature in the presence of low oxygen concentrations by the reaction 2 Si + O₂ → 2 SiO. This reaction generates a product—SiO—which is volatile at sufficiently high temperatures. Since atomic-scale etching is random, the removal of Si atoms by this reaction will lead to roughening of the surface (provided temperatures are not so high as to initiate surface smoothing by Si surface self-diffusion). The regime in which oxygen concentrations are sufficiently low that etching dominates over oxidation is a reaction domain made readily accessible for the first time by the ultraclean, integrated processing environment, while this etching regime cannot usually be encountered in conventional MOS processing.

A crucial aspect of the chemical behavior of MOS breakdown can be assessed from the data shown in Fig. 1. Here MOS capacitors were fabricated by an integrated processing involving the pre-oxidation and thermal oxidation, followed by ex situ Al gate deposition. As mentioned above, a final HF dip cleaning leads to low field breakdown if temperature ramp-up is carried out in 1 atm. pure Ar. To confirm the chemical mechanism responsible, in these experiments a pre-anneal in 1 atm. pured Ar prior to thermal oxidation was added, followed by ramp-up in 1 atm. oxygen. Since oxygen ramp-up without the Ar pre-anneal yields high breakdown fields, modifications of the Ar pre-anneal could be employed to assess mechanistic behavior.

The Ar pre-anneals were carried out at different temperatures and in the presence of controlled, known amounts of O₂ added to the 1 atm. purified Ar ambient. The data points shown in Fig. 1 lie at specific pre-anneal temperatures and O₂ concentrations (in 1 atm. Ar) investigated, and the resulting average breakdown fields for the corresponding MOS capacitors are indicated as numbers next to the data points. In Fig. 1 the vertical axis is labelled in terms of both the O₂ partial pressure (in torr) and concentration in the 1 atm. Ar (in ppm).

The average breakdown fields for the data points display a clear demarkation between high and low fields, as indicated by the sloping line in the figure. This line represents a kind of "reaction boundary"
between good and poor breakdown behavior. Not that the plot in Fig. 1 is presented in a form similar to that of an Arrhenius plot for chemical kinetics, and the slope of the reaction boundary might be taken to correspond in this sense to an "activation energy". This slope of the boundary in Fig. 1 suggests an activation energy of \( \sim 3.9 \) eV, essentially the same as that for SiO desorption [8], [9].

This result is meaningful in two ways. First, prior evidence has suggested strongly that the chemical systematics of breakdown, as related to the presence or absence of an oxide passivation layer on the surface before ramp-up and oxidation, is based on a mechanism in which the oxygen etching reaction 2 Si + O \(_2\) \(\rightarrow\) 2 SiO leads to desorption of a volatile SiO product. Second, Smith and Ghidini determined the desorption activation energy of SiO specifically by identifying the reaction boundary between (i) oxidation of the Si surface at higher O\(_2\) or H\(_2\)O partial pressures and (ii) etching at lower partial pressures, precisely the kind of behavior postulated to explain the chemical systematics of electrical breakdown reported here. With the Ar/O\(_2\) pre-anneal cycle imposed here specifically to provide a time during which the oxidation vs. etching response of the surface may be tested, the results in Fig. 1 demonstrate quite conclusively that the Si surface etching in low concentrations of oxygen is the mechanism responsible for the breakdown behavior observed. Thus the present work demonstrates that the reaction of Si etching in low oxygen concentrations—earlier elucidated as a surface science phenomenon by Smith and Ghidini—has profound implications for manufacturing technology employing advanced processing equipment and approaches.

While Al-gate MOS capacitors are convenient test devices for mechanistic relevance to demonstrate that the same overall chemical behavior occurs for polySi gate structures as well. Breakdown distributions for polySi gate MOS capacitors are shown in Fig. 2, where the polySi layer was grown in the hot wall reactor after completion of the thermal oxidation process. For ramp-up in Ar after HF dip cleaning, low field breakdown is observed, while for ramp-up in O\(_2\) high breakdown fields are found. (The low field breakdown events are believed a consequence of particulate contamination, since the wafers were moved through normal laboratories—not a clean room—from the cluster tool to the laboratory for further processing.) The comparison between Ar and O\(_2\) ramp-up mirrors the behavior found previously for Al-gate capacitors, and is consistent with the surface etching model explained above.

IV. IMPLICATIONS FOR INTEGRATED PROCESSING IN MANUFACTURING

The implications of this etching mechanism are represented schematically in Fig. 3, which outlines process sequences in conventional and integrated MOS gate dielectric processing in terms of the temperature and O\(_2\) partial pressure to which the Si surface is exposed during processing. The conditions of room temperature and high temperature (e.g., that for thermal oxidation) are represented on the inverse temperature horizontal axis, while the oxygen partial pressure is given on vertical log scale from negligible oxygen concentration (purified Ar or vacuum) to 1 atm O\(_2\). In conventional MOS processing (a), the pre-oxidation clean causes some chemical oxidation of the surface; then the wafer is removed into air (20% O\(_2\), plus H\(_2\)O), put onto the furnace cantilever, and moved into the oxidation furnace (1 atm O\(_2\)). As the wafer is heated it is continuously exposed to large quantities of oxygen. Surface etching by oxygen cannot occur, for two reasons. First, the chemical oxide layer remaining from standard RCA clean passivates the surface against etching by trace oxygen (it forms an etch mask). Second,
the ambient conditions at the surface never approach the etching regime, but rather remain well within the oxidation regime at all times. This explains why etching and roughening by oxygen species is not observed, and indeed is not accessible, in conventional MOS processing.

Integrated processing in ultraclean tool environment makes possible surface etching by oxygen for the first time, as depicted in Fig. 3(b). Following conventional RCA cleaning, the final pre-oxidation cleaning step of HF dipping removes all passivating oxide from the Si surface, leaving only a rather weakly bonded hydrogen monolayer and a lesser amount of fluorine [10], [11]. Upon removal from the HF in a purified Ar glove box, the surface is not exposed to oxygen, and transfer in the vacuum system to the oxidation reactor under ultrahigh vacuum conditions presents negligible levels of oxygen exposure. When the wafer is then placed in the oxidation reactor and ramped up in temperature to oxidation temperatures under 1 atm. Ar ambient, the hydrogen desorbs from the surface (by 550°C)13,15 and the Si surface moves into the etching regime of the reaction diagram; at this point whatever trace concentrations of oxygen (or water) are present as impurities will etch the surface. Furthermore, once oxidation temperature is reached and O2 is admitted to displace the Ar, a transient etching effect will occur while the oxygen concentration builds up.

Clearly, the amount of etching and roughening which takes place in this situation depends on concentrations and kinetics. If the time spent in the etching regime is sufficiently short, or the oxygen concentration sufficiently low, etching may be negligible. The closer the system is to the reaction boundary — on the etching side — the more extensive should be the etching and roughening; however, we know of no quantitative assessment of roughening rates and their relationship to breakdown properties to date, and the breakdown data in Fig. 1 are not sufficient to show a dependence on proximity to the reaction boundary from the etching side. Finally, it should be recognized that subtle physical and chemical phenomena on an atomic scale will modulate the extent and spatial distribution of Si surface roughening. Statistical fluctuations will produce at the same time greater and lesser amounts of etching and oxidation at different points on the surface. Preferential reaction at surface atomic steps [16] will modify the spatial pattern of etching. And when pressure/temperature conditions are near the reaction boundary between etching and oxidation, local regions may be lightly oxidized and thereby passivate the surface to prevent etching, leading to micromasking which is evidenced as even more dramatic roughening of the surface (as confirmed by ~ 70Å roughness in scanning tunneling microscopy images4). Because of these local effects which occur across a large span of wafer surface, the reaction boundary is in some sense a graded or fuzzy transition.

Two approaches to prevent etching-induced roughness are illustrated schematically in Fig. 3(c) and 3(d). In (c) O2 is admitted during temperature ramp-up in sufficient quantity to maintain the surface on the oxidation side of the reaction diagram, thus forming a thin passivating oxide layer immediately to resist etching and ensuring that any unoxidized portions of the surface are on the oxidation side of the reaction boundary. The simplest embodiment of this prescription is to ramp-up in 1 atm. O2, as in conventional MOS processing.

In (d) the HF dip is followed by a UV-ozone treatment in the glove box to intentionally grow a thin chemical oxide on the Si surface after HF dip11,11 which passesivates the surface against the etching reaction. Once this passivating oxide layer is in place, the recipe for ambient during ramp-up is not critical (providing temperatures are not reached which would initiate decomposition and removal of the oxide by the interfacial decomposition reaction17,18 Si + SiO2 \(\Rightarrow\) 2 SiO). With the etching/roughening mechanism identified, the prescriptions for cluster process integration represented in Fig. 3(c) and 3(d) show how to prevent etching and roughening of the surface. Although one might argue that the new chemical regime accessible by ultraclean integrated processing has here generated new and intrinsic problems, the mechanistic insights achieved lead to simple prescriptions for cluster process integration, so that other expected advantages of cluster processing can be realized.

It is interesting to note a closely related phenomenon which occurs in post-oxidation annealing of thermal oxide structures. If annealing is carried out in oxygen-deficient ambients at > 900°C, interfacial decomposition leads to oxide and Si removal as volatile SiO17,19. At lower temperatures (> 750°C) the precursor to this reaction produces electrical active defects as seen in breakdown and hole trapping properties. The dependence of breakdown on oxygen concentration and temperature is virtually the same as that seen in Fig. 1, demonstrating that fundamental Si-O chemistry is producing deleterious defect reactions. Again the solution is to assure a sufficient though low concentration of oxygen is always present at high temperatures to prevent the decomposition reaction when ultraclean processing environments are available.

Water vapor is a dominant impurity in reactors or vacuum systems (unless baked to achieve ultrahigh vacuum), so that H2O is the oxygen species of greatest concern in processing ambients. Smith and Ghidini8,9 found the role of O2 and H2O to be very similar in causing Si surface etching, so the results presented here can be taken as representative of the behavior of low H2O concentrations.

Although in retrospect the ramp-up in Ar seems not a good choice, its original motivation was not difficult to understand. The clean Si surface would seem an ideal starting point, being well-defined and offering the possibility to control the oxidation time and thickness by starting the process at just the time when O2 is admitted to displace the Ar. The results here show that the formation of a thin oxide passivation layer, e.g., by UV-ozone treatment, is essential to prevent deleterious etching of the Si surface by trace oxygen-containing impurities. Indeed, this approach has recently been utilized20 for ultraclean processing to form ultrathin thermal oxide films.

It is tempting to conclude that ultraclean conditions — a goal of advanced process and tool technology — are detrimental, at least in certain cases. However, manufacturing effectiveness is not well served by irreproducibility, and less clear conditions typically entail unpredictable concentrations of other impurities (e.g., organics), whose chemistry in these processes and process sequences may well be deleterious. Highly reproducible manufacturing processes would seem more likely when ambient conditions are highly controlled, and when important chemical reaction pathways (like etching and decomposition here) are identified and intentionally controlled.

The value of a surface oxide passivation layer to integrated MOS processing bears significance for tooling issues. It suggests that the pre-oxidation clean, if a wet process primarily compatible with atmospheric pressure conditions, could be integrated effectively with atmospheric pressure thermal oxidation in a cluster tool which is maintained only at atmospheric pressure, with no vacuum capability. However, that leaves an open question in regard to the integration of the polySi process and tool, which is usually practiced as a low pressure CVD process.

**Acknowledgment**

The authors are grateful to Srinandan Kasi for his contributions to the surface cleaning procedures used here and to R. McIntosh for assistance in wafer processing. They also appreciate the help of E. M. Kummer and J. E. Pucykowski in fabrication and maintenance of the oxidation reactor.
REFERENCES


