Ion beams in silicon processing and characterization

E. Chason and S. T. Picraux
Sandia National Laboratories, Post Office Box 5800, Albuquerque, New Mexico 87185

J. M. Poate
New Jersey Institute of Technology, College of Science and Liberal Arts, 504 Cullimore Hall,
University Heights, Newark, New Jersey 07102

J. O. Borland
Genus Corporation, 4 Muliken Way, Newburyport, Massachusetts 01950

M. I. Current
Applied Materials, Implant Division, 9020-I Capital of Texas Highway North, Austin, Texas 78759

T. Diaz de la Rubia
Lawrence Livermore National Laboratories, Post Office Box 808, Livermore, California 94550

D. J. Eaglesham
Bell Laboratories, Lucent Technologies, 700 Mountain Avenue, Murray Hill, New Jersey 07974-0636

O. W. Holland
Oak Ridge National Laboratory, Post Office Box 2008, Oak Ridge, Tennessee 37831-6057

M. E. Law
University of Florida, Department of Electrical Engineering, 339 Larsen Hall, Gainesville,
Florida 32611-6200

C. W. Magee
Evans East Incorporated, 666 Plainsboro Road, Plainsboro, New Jersey 08536

J. W. Mayer
Arizona State University, Center for Solid State Science, Tempe, Arizona 85287-1704

J. Melngailis
University of Maryland, Energy Research Building, College Park, Maryland 20742-3511

A. F. Tasch
University of Texas, BRC-MER Building, Austin, Texas 78712

(Received 12 June 1996; accepted for publication 23 January 1997)

General trends in integrated circuit technology toward smaller device dimensions, lower thermal budgets, and simplified processing steps present severe physical and engineering challenges to ion implantation. These challenges, together with the need for physically based models at exceedingly small dimensions, are leading to a new level of understanding of fundamental defect science in Si. In this article, we review the current status and future trends in ion implantation of Si at low and high energies with particular emphasis on areas where recent advances have been made and where further understanding is needed. Particularly interesting are the emerging approaches to defect and dopant distribution modeling, transient enhanced diffusion, high energy implantation and defect accumulation, and metal impurity gettering. Developments in the use of ion beams for analysis indicate much progress has been made in one-dimensional analysis, but that severe challenges for two-dimensional characterization remain. The breadth of ion beams in the semiconductor industry is illustrated by the successful use of focused beams for machining and repair, and the development of ion-based lithographic systems. This suite of ion beam processing, modeling, and analysis techniques will be explored both from the perspective of the emerging science issues and from the technological challenges. © 1997 American Institute of Physics.
I. INTRODUCTION

Ion implantation is one of the most important processing tools in Si integrated circuit technology. Its discovery by Ohl at Bell Laboratories in 1952 and subsequent development in industrial, governmental, and university laboratories to become a workhorse of the industry makes fascinating reading. Today ion beams are ubiquitous in Si technology (see Fig. 1). There has been the recent realization, however, that the move to smaller dimensions in integrated circuit technology is confronting ion implantation with a new round of physical and engineering challenges. The physical challenges center around the interaction between the defects and dopants introduced during implantation. When device dimensions were in the micrometer range, understanding these interactions was not crucial. Now, with device dimensions well into the submicrometer range, dopant-defect interactions loom large and quantitative information is needed.

This review will show that on the scientific level we are entering a new era of quantifying defect production in Si and of gaining a microscopic understanding of dopant-defect interactions. The engineering challenges center around the development of new generations of low and high energy implantation machines to accommodate the many implantation steps used in the fabrication of a single integrated circuit. We primarily focus on energies greater than 1 keV and do not review here the related technologies of plasma-enhanced etching and deposition. Focused ion beams are having considerable impact in diverse areas of semiconductor technology and these will be reviewed. The technology that is facing the most severe challenge is that of lithography where optical techniques dominate. Eventually shorter wavelength beams
will be required for the fabrication of patterns and the status of ion beams for lithography will be discussed.

The progress of ion implantation technology has been well documented in a series of books 2–7 and in major international conference proceedings 8–11 over the past twenty years. During that time remarkable progress has been made in understanding the particle-solid collision phenomena so that range profiles of implanted dopants can be calculated with considerable accuracy. The furnace annealing conditions to remove the damage introduced by the implantation and leave the implanted dopants on lattice sites, and fully electrically active, have also been well explored. The basic understanding of the processes involving the formation and removal of point defects has also progressed but is not so well advanced as that for range calculations. While these basic advances were occurring, the industry was also advancing at a remarkable rate as shown by a type of Moore plot in Fig. 2. Figure 2 shows that the amount of information that can be stored on a Si integrated circuit has increased exponentially with time. The relative information content of a page, book, and encyclopedia is shown for comparison.

Implantation protocols have become quite complex to meet the demands of small device structures such as the typical 0.35 μm gate length Complementary MOS (CMOS) transistor shown in Fig. 3. Although it is beyond the scope of this review to detail all the electronic engineering associated with dopant incorporation and the concomitant carrier and field distributions, Fig. 3 illustrates the sophistication of current implantation technology. As many as 20 different implantations covering a wide range of energies (5 keV–1 MeV), doses (10^{11}–10^{15} cm^{-2}) and dopant species (B, P, As) are typically used to tailor dopant profiles. The lowest energies are for junction formation and the highest for substrate or well doping. In recent years a fascinating physical phenomenon critically important to the formation of shallow dopant profiles has emerged. The data of Michel et al. 13 (Fig. 4) illustrates this phenomenon. The B implantation results in the usual Gaussian shaped profile, but on annealing a burst of diffusion is observed which saturates with time. The magnitude of this diffusion is many orders of magnitude greater than equilibrium B diffusion in Si and extends to a depth of approximately 0.1 μm. This effect was of little importance when the critical device dimensions were a micrometer or greater, but such diffusion lengths can be critical for devices in the submicrometer regime. It is caused by the release of Si interstitials from the implantation damage and their subsequent interaction with the B atoms. This transient enhanced diffusion of B poses severe challenges to both the implantation process engineer and the device modeler.

**TABLE I. Projected MOS source/drain doping technology requirements (from Ref. 12).**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Transistor gate length (μm)</td>
<td>0.35</td>
<td>0.25</td>
<td>0.18</td>
<td>0.13</td>
<td>0.10</td>
<td>0.07</td>
</tr>
<tr>
<td>X_i at channel (nm)</td>
<td>70–150</td>
<td>50–120</td>
<td>30–80</td>
<td>20–60</td>
<td>15–45</td>
<td>10–30</td>
</tr>
<tr>
<td>Surface concentration (cm^{-2})</td>
<td>10^{19}</td>
<td>10^{18}</td>
<td>10^{19}</td>
<td>10^{19}</td>
<td>10^{19}</td>
<td>10^{19}</td>
</tr>
<tr>
<td>Wafer diameter (mm)</td>
<td>200</td>
<td>200</td>
<td>300</td>
<td>300</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>Metal impurity (at. cm^{-2})</td>
<td>5×10^{10}</td>
<td>2.5×10^{10}</td>
<td>1.0×10^{10}</td>
<td>5×10^{9}</td>
<td>2.5×10^{9}</td>
<td>&lt;2.5×10^{9}</td>
</tr>
</tbody>
</table>

FIG. 2. The amount of information that can be stored on a Si integrated circuit has increased exponentially with time. The relative information content of a page, book, and encyclopedia is shown for comparison.
The implantation process is one of considerable physical complexity and much remains to achieve a full microscopic understanding. This review will detail some of the current research where theoretical and experimental advances are helping to quantify the phenomena. Let us consider the case of a 5 keV B ion implanted into Si. The ion will stop in a fraction of a picosecond creating a damage cascade. The Si lattice is left in much disorder with a high density of Frenkel pairs—interstitials and vacancies. On annealing most of these pairs will recombine but some Si interstitials will agglomerate into clusters and then be re-emitted during later high temperature annealing. Experiments and theory are now beginning to provide a quantitative understanding of this defect formation and emission process. Once the Si interstitial is released, it can diffuse through the lattice and interact with a B atom causing the ejection and subsequent diffusion by an interstitialcy mechanism. Although the exact atomistic nature of the interstitialcy mechanism is still not known, the B atom eventually comes to rest on a lattice site and becomes electrically active as a p-type dopant.

The intent of this review is to detail the status of the ion beam field in Si processing and indicate where the challenges are. Many aspects are now undergoing a renaissance in understanding as a result of questions that have been raised by the severe engineering constraints imposed by shrinking device dimensions. We first discuss the area of dopant profile modeling before any annealing is carried out. The use of computationally efficient empirical models and Monte Carlo techniques to simulate these profiles in the presence of channeling are reviewed.Implant-induced damage production has a critical effect on dopant mobility during annealing, so atomistic Monte Carlo and molecular dynamics simulations of damage accumulation are also discussed.

Once the implant has been performed, the Si wafer must be annealed to activate the dopants. Understanding defect formation and annealing for planar integrated circuits in the submicrometer regime is one of the most challenging problems in radiation damage studies. The atomic scale dimension, with the presence of surface and interface sinks, demands precise understanding of the basic mechanisms of defect migration and recombination. Moreover, the device engineers require accurate estimates for the formation and ultimate fate of both point and extended defects. While considerable understanding was achieved early on in the production of defects and of vacancylike defect evolution, the release of defects at intermediate temperatures (400–700 °C) and the role of interstitials remained largely a mystery. This regime is particularly important in understanding transient enhanced diffusion, a major concern in achieving even tighter dimensional control of dopants. Much progress in this area has been made in recent years by developments in quantitative transmission electron microscopy (TEM) coupled with detailed diffusion studies and with advances in computer modeling of the cascade and damage processes.

The use of high energy and high dose beams are now becoming important in Si technology. The high energy implants are being used to directly fabricate doped well or tub structures in which the MOS transistors are fabricated. The conventional approach is to form the wells by low energy implantation followed by high temperature and lengthy anneals. High energy implantation does this directly thus removing several fabrication steps and time consuming anneals. Coupled with the interest in high energy implantation is the need to understand gettering phenomena at the atomistic level. Table I shows the very low level of metal impurities tolerable in current and future devices. Several implantation scenarios are now being investigated to getter impurities from the active regions of devices. Two areas, void formation by inert gas implantation and high energy implantation, are giving fundamental information on gettering mechanisms. Their use as potential gettering agents will be reviewed. Very high dose beams of oxygen can be used to form continuous and buried SiO2 layers in Si. This technology is a promising candidate for dielectrically isolated Si devices and its utilization requires understanding of defects in regimes very different from the lower doses required for dopant incorporation.

Modeling and understanding of defect and dopant processes are ultimately constrained by our ability to measure implant profiles on the scale necessary for current device technology. In particular, the understanding of lateral dopant profiles is limited by the inability to measure two-dimensional (2D) implantation profiles in the submicrometer regime. We discuss the status of current techniques based on secondary ion mass spectrometry (SIMS) and ion beam analysis for achieving the necessary resolution to characterize dopant level concentrations with the spatial resolution required by device dimensions.

The coupling between research into the physics and materials science of ion beam processing and the development of new machines is an important part of ion processing technologies. We complete this review by detailing some of the recent progress in machines and the developments that are required for future technologies. The issues span the very tight constraints needed for implantation across large wafers to the demand for focused beams for micromachining and analysis. Indeed, ion beam technologies are now being explored for the next generation lithographic systems.
II. SHALLOW IMPLANTATION: PROFILES, DEFECTS, AND DOPANT DIFFUSION

Ion implantation has been used in Si MOS integrated circuit fabrication since the early to mid-1970s for the purpose of forming the doping profiles required in MOS transistors and memory cells. Its advantages over other doping methods include accurate dose control, higher purity of the dopant species, reproducibility of the impurity profiles, lower process temperature (however, electrical activation at higher temperatures is still required), the ability to selectively dope the Si, and the ability to tailor the doping profile. Ion implantation continues to be very widely used today and is expected to remain as the dominant doping method in the foreseeable future in the fabrication of integrated circuits.

As device dimensions decrease, scaling of the doping parameters must occur concurrently in order to achieve and maintain adequate device electrical characteristics such as good turnoff, high drive current, and high reliability. This scaling requires increasingly compact (shallow and rapidly varying) doping profiles in both the source/drain junctions and the channel regions of the MOS transistor. These shallow doping profiles also require much lower thermal budgets than previous generations. In the past, thermal budgets were sufficiently large so that the final dopant profile did not depend appreciably on the implant parameters. For shallow profiles, the final profile depends much more on the implant conditions, and it is well established that the implant-induced damage affects the final (after thermal processing) doping profile due to defect enhanced diffusion.

Economic and technological factors have resulted in tremendous emphasis on rapid technology development and to a large increase in the use of technology computer-aided design (TCAD) in the simulation of the MOS device and the fabrication process. This, in turn, has generated a demand for computationally efficient, comprehensive, and validated process models, including models for the simulation of the as-implanted impurity and damage profiles. Process modeling and control of dopant profiles on these new length scales require greater understanding of implantation profiles and defect production than has been needed before. In this section we review the status of current programs to model implantation profiles, defect production, and the evolution of dopant profiles during processing. We first discuss recent progress in the use of efficient phenomenological approaches as well as physically based models for predicting as-implanted profiles. This is followed by a discussion of molecular dynamics calculations of damage production and a novel approach to quantifying ion-induced defect production using surface trapping. The evolution of dopant profiles and the interactions of dopants with ion-induced defects, discussed in the final part of this section, are especially important for the production of shallow implant profiles. Recent progress using quantitative TEM and dopant superlattices is expanding our understanding of defect-enhanced dopant diffusion. Attempts to analytically and atomistically model this complex process are discussed, as well as recent first-principle calculations that are shedding light on atomic-level defect-dopant interaction mechanisms.

A. As-implanted dopant profiles

Understanding the as-implanted dopant profile is a critical first step in controlling shallow doping layers. Because the ion implant is performed in single crystal Si, the as-implanted profile has a strong dependence on implant angle, dose, and substrate temperature, in addition to energy. Calculational approaches are based on both semi-empirical methods and physically based models. In the semi-empirical approach, a large amount of experimental data is used to refine model parameters and their dependence on input parameters. These models have the advantage of being computationally efficient. Physically based models rely on a description of the basic physical processes controlling the system and help to validate the physical understanding of the processes involved.

The Dual–Pearson fitting procedure has been highly successful in the semi-empirical model development for B,
BF$_2$, As, and P implants$^{15-18}$ because of its ability to account so well for the dependence of both the randomly scattered and the channeled parts of the implanted profile on all of the key implant parameters. These models have explicit dependence on energy, dose, tilt angle, and rotation angle, and examples of their excellent predictive capabilities are illustrated in Fig. 5. This analytic function approach is much more computationally efficient than physically based models, and has been demonstrated to accurately describe the impurity profile over four to five orders of magnitude variation in the concentration$^{14-18}$ In addition, this approach has been demonstrated to model the wide range of different shaped profiles resulting from channeled implants, off-axis (minimum channeling) implants, different doses, and low mass (B) and high mass (As) implant species.$^{14-18}$ The development of accurate models of this type, however, requires extensive experimental data, especially in order to correctly understand and predict the detailed channeling dependence of the profile on the various parameters. Towards this end a very large number of wafers has been implanted, and over 2000 SIMS depth profiles have been carefully measured in order to determine values, and an interpolation scheme for the nine parameters required by the Dual–Pearson model.

Physical models are typically based on Monte Carlo computer simulations using the binary collision approximation (BCA). BCA simulations depend on calculating the classical scattering integral for two particles, one moving and the other stationary, that interact through a screened Coulomb interaction potential. Atoms are then set in motion according to the kinematics of the collision and go on to interact with other atoms in the system, but only while their energy exceeds a threshold value, usually set at the minimum energy required to produce a Frenkel pair in the material ($5\text{ eV}$ for Si). Programs of this type include MARLOWE, originally developed at Oak Ridge National Laboratory to study radiation effects in crystalline materials$^{19}$ and UT-MARLOWE, an enhanced version developed at the University of Texas at Austin for studying implantation in Si. Inherently a three-dimensional (3D) simulator, UT-MARLOWE has been demonstrated to accurately predict as-implanted impurity profiles for B ($5-80$ keV), BF$_2$ ($5-65$ keV), and As ($5-180$ keV) implants into single-crystal Si.$^{20-22}$ Ziegler–Biersack–Littmark (ZBL) pair-specific (e.g., As–Si, B–Si, O–Si, Si–Si, etc) interatomic potentials have been implemented in the nuclear scattering BCA model in order to improve the accuracy of the model.$^{23}$ The ion beam divergence, native oxide layers, and wafer temperature during implantation are taken into account. In addition, new models for the electronic stopping and dose dependence (cumulative damage) have recently been developed.$^{20-22,24,25}$ The electron stopping models have explicit dependence on the electron concentration in the Si lattice as determined from ab initio calculations.$^{26}$ These models have resulted in excellent predictive capabilities for the detailed dependence of arsenic and boron as-implanted profiles on all key implant parameters for B, BF$_2$, and As implants into single-crystal Si, partially damaged (due to ion implants) single-crystal Si, and amorphous Si.$^{20-22}$ This approach is illustrated in Fig. 6 which demonstrates the good agreement between the simulated profiles and the experimental data for different implant parameter variations.

The UT-MARLOWE model for boron has been extended to predict the profiles for boron implants through overlying oxide layers into single-crystal Si, since this situation commonly occurs in integrated circuit fabrication. In addition, the cumulative damage models for B, BF$_2$, and As in UT-MARLOWE produce reasonable estimates of the Si interstitial and vacancy distributions resulting from the implant-induced damage. The cumulative damage models take into account nearest neighbor recombination of interstitials during the implant, the presence of highly localized disordered regions (also referred to as “amorphous pockets”), and a threshold concentration of interstitials and vacancies for amorphization. A threshold of 10% displacement of the Si lattice atoms is used as the criterion.$^{27}$ These cumulative damage models not only accurately predict the dose dependence of the impurity profiles, but also predict the experimentally measured amorphous layer thicknesses.

FIG. 6. Comparison of the impurity profiles predicted by UT-MARLOWE with experimentally measured profiles. (a) Dependence on implant dose for BF$_2$ implants at 35 keV. (b) Dependence on tilt angle for As implants at 180 keV (from Ref. 22).
A critical need for two- and three-dimensional implant models has also developed due to the very small device sizes and the dense packing of these devices in integrated circuits. Also, effects due to implants through overlying oxide layers, the orientation of the masking edge, and the masking layer thickness are of major importance. It is highly desirable to have two- and three-dimensional models that can accurately predict the detailed profile dependence on energy, dose, tilt angle, rotation angle, masking layer thickness (to account for implant shadowing at the masking layer edge in addition to correctly accounting for the implant profile dependence on overlying layer thickness), and masking edge orientation.

To begin to address these needs, a 2D model for B implants into (100) single-crystal Si through overlying oxide layers was developed. The 2D implant profile uses a semi-empirical equation based on the Pearson functions. Since 2D characterization techniques are not capable of measuring numerous profiles over a large dynamic range with a resolution of 2–5 nm, the results of this model were compared with profiles generated by UT-MARLOWE Monte Carlo simulations. The distributions generated by this equation closely match the simulated distributions, as shown in Fig. 7.

B. Damage production

1. Molecular dynamics calculations

A powerful method to study ion–solid interactions that does not make use of the BCA is molecular dynamics computer (MD) simulation. Classical MD treats a system of interacting degrees of freedom (atoms) by solving the classical equations of motion iteratively to obtain their phase space trajectories. Because of its inherent simplicity, MD is an extremely powerful tool to obtain atomic scale information and physical insight into the mechanisms of mass transport and microstructural evolution in materials subjected to external, nonequilibrium driving forces. In particular, because of the length scale (e.g., 300 Å × 300 Å × 300 Å cubes containing 10^6 atoms) and time scale (~10^-9 s) accessible to the simulations, MD is ideally suited to study the energy deposition and subsequent thermalization of the cascade region, and therefore provides atomistic details of the mechanisms of defect production and phase transformation during ion beam processing of semiconductors.

At present, MD simulations are being used to study semiconductor processes such as low energy reactive ion etching, sputtering, and ion implantation for shallow junction technology. Because of the short times accessible, long range diffusion of dopants and defects cannot be modeled with MD. However, as discussed below, kinetic Monte Carlo (MC) methods, which have been used before to study adatom and vacancy diffusion on Si surfaces and to model defect evolution in metals during high temperature neutron irradiation, can be used to couple the MD simulation to the much longer time regime of damage accumulation and dopant diffusion.

The simulations discussed below were performed with the Stillinger–Weber (SW) potential for Si (Ref. 36) using a MD code optimized for ion–solid interaction studies (MOLDYCAST). Although many other potentials for Si exist in the literature, experience with the SW potential indicates that it is a reasonable representation of Si for the study of ion beam processing, accurately describing many properties of small Si clusters and bulk Si (including melting and rapid recrystallization kinetics). Comparison to very recent density functional calculations with ab initio pseudopotentials shows very good agreement between these electronic structure methods and the simple SW potential. The calculations described here were performed on the 256 processor Cray Research Inc. T3D MPP computer at Lawrence Livermore National Laboratory.

To simulate an ion beam interacting with a surface, a system of atoms with the desired crystal structure is generated with periodic boundary conditions and allowed to come to equilibrium with a thermal reservoir. After equilibration of the system, one atom [the primary knock-on atom (PKA)] is given the velocity corresponding to the implantation energy and beam angle of incidence that needs to be simulated. The ion trajectory, as well as that of all subsequent secondary and higher energy recoils, is followed and the simulation proceeds until the displacement cascade region reaches equilibrium with the surrounding thermal bath, i.e., until all deposited energy has dissipated into the boundary and thermal reservoirs. At that point, a new PKA can be launched into the simulation box and the whole process is repeated. In this manner, dose can be accumulated, and effects such as spatial and temporal cascade overlap can be easily studied.

For the simulation of ion beam processing the model interatomic potential has to be modified in the region of short interatomic separation. This is because many atoms interact at spacings well inside the core–core overlap region during the cascade process. This modification requires that some information regarding the form of the potential be known for distances of approach shorter than those characteristic of, for example, the pair distance in a (110) dumbbell. Presently used potentials have been found to be in agreement with experimental values of the orientation dependence of $E_d$. These results indicate that these potentials are well

![Fig. 7. Illustration of the predictive capability of the computationally efficient 2D model for boron implants. The implant is performed through a 0.4-μm-wide window in the mask at an energy of 80 keV, a dose of 1 × 10^13 cm^-2, and tilt and rotation angles of 10° and 0°, respectively. The 2D profile is along the (110) direction. The solid isodensity curves are generated by UT-MARLOWE, and the dotted isodensity curves are predicted by the new 2D model (from Ref. 28).](Image)
suited to describe the low energy defect production mechanisms. The model potential is then modified for energies higher, or distances of closest approach in a two-body collision shorter, than those characteristic of these threshold energy events. One typical solution in Si (Ref. 44) is to spline the two-body part of the SW potential to the so-called Universal, or ZBL, potential.45 The ZBL potential is a screened Coulomb interaction potential designed to describe properly the scattering properties of a very wide range of two-atom systems, and is widely used in BCA codes such as the aforementioned MARLOWE and TRIM.

Results from a typical simulation of a displacement cascade for a 5 keV Si ion impacts on Si are illustrated in Fig. 8.32 Only atoms with potential energies $E_p > 0.2$ eV above the ground state are illustrated. The diameter provides an indication of atom potential energies in the range of 0.2–1 eV or larger. The shade of the spheres indicates atomic level stress, $\sigma$, where $\sigma=\Omega^{-1} \frac{\partial V}{\partial \zeta}$, $\Omega$ is the atomic volume of the crystal, and $\zeta$ is a distance scaling factor. The shade scale ranges from very dark for tensile stress to light gray for compressive stress. The initial substrate temperature was 80 K in this case.

At the start of the collision process, the energy is concentrated in a relatively small number of atoms with large potential energies, as shown in Fig. 8(a). At this time, 0.1 ps after the start of the 5 keV atom, the region of damage has not extended to its full depth, and the cascade region is under large compressive stress. The maximum number of atoms with $E_p > 0.2$ eV occurs about 0.2 ps after the first collision; at this point the average total energy per atom is $\approx 1$ eV for those atoms with $E_p > 0.2$ eV. The average potential energy of the particles illustrated in Fig. 8(b) ($t=1$ ps) is 0.47 eV whereas the latent heat of fusion of the Si model is 0.325, suggesting that these particles may have properties similar to that of liquid Si. Figure 8(c) shows that about $5 \times 10^2$ atoms still have $E_p > 0.2$ eV after 8 ps; the average potential energy of these atoms is 0.34 eV.

One of the interesting observations that may be derived from these simulations is the fact that very few isolated Frenkel pairs are produced by these cascades. This can be understood by considering that isolated point defects are produced in cascades by replacement collision sequences (RCSs) along low index crystallographic directions.46,47 As we discussed above, such RCSs in Si are extremely short, two to three atomic replacements long at most.43 Therefore, the probability of such a RCS producing a stable (i.e., separated beyond its own spontaneous recombination volume) vacancy–interstitial pair during a keV cascade is very small. On the other hand, large pockets of unrelaxed amorphous material are produced by the displacement cascades. For these 5 keV cascades, the disordered regions contain an average of $\sim 800$ atoms, i.e., seven times the number of displaced atoms predicted by the BCA calculations. Because of the energetic collisions and the density changes induced by the incoming particle, these amorphous regions frequently have densities that differ from that of the crystal. Also, because of their large surface-to-volume ratio and the fact that they are surrounded by crystalline material, these amorphous pockets are highly unstable, and recrystallize at much lower temperatures than a stable planar amorphous/crystal interface. Upon annealing for several picoseconds at elevated temperature, such density fluctuations result in the appearance of vacancies and self-interstitial atoms in the recrystallized material. These defects can then participate in the transient enhanced diffusion of dopant atoms that is commonly observed in low energy ion implanted Si. Note that out of approximately 500 atoms originally present in a typical amorphous pocket, only four or five defects at most remain as damage.32

2. Quantifying defect production by surface trapping

Although the production and diffusion of ion induced defects are critical to understanding their effect on dopant distributions, the transient nature of these defects makes it difficult to measure their production directly. Quantitative TEM has made significant progress in explaining transient enhanced diffusion by quantifying the number of ion-induced interstitials trapped in extended defects (see Sec. II C). Another method that was recently developed takes advantage of the fact that defect diffusion is slower on the surface of semiconductors than in the bulk so defects that reach the surface are trapped there. The process is shown schematically in Fig. 9(a): defects are created in the bulk and then diffuse to the surface where they are trapped by their reduced mobility at the surface. Probes of surface morphology, such as reflection high energy electron diffraction (RHEED)49 and scanning tunneling microscopy (STM),49 can then be used to determine the number of defects produced per ion. Because RHEED is nondestructive and can be performed in real time, it can also be used to study the interaction among vacancy and adatom type defects on the surface in order to study the kinetics of surface defect recombination.33,49,50

Up to this point, RHEED was used primarily to study the defect yield for very low energy ions in the range of 70–500
eV to enhance surface processing. By measuring the rate at which the surface roughens during ion bombardment, the number of surface defects created per incident ion can be determined. Measurements of the surface defect yield for several ion species (He, Ar, and Xe) on the Ge(001) surface are shown in Fig. 9(b). The measurements were performed at −100 °C in order to freeze out defect diffusion on the surface and to prevent adatom and surface vacancy recombination. The effect of temperature on the surface defect yield is shown in Fig. 10 for Xe and Ar ions. At low temperature, the defect yield is approximately 12 defects per ion, in reasonable agreement with TRIM Monte Carlo simulations of defect production using the BCA, indicating that roughly all of the defects created by the ion migrate to the surface. At higher temperature, the defect yield decreases rapidly to approximately one defect per ion, which is the value of the sputter yield at this energy. This rapid decrease in the defect yield is attributed to the increase in surface diffusion with temperature that allows the defects on the surface to become mobile and recombine. Only the uncompensated sputter vacancy remains after the surface recombination process. Monte Carlo simulations using defect diffusion energies obtained from previous measurements show that these results are consistent with the mechanism of surface recombination.

It is important to note that RHEED is only sensitive to surface roughness and not to bulk defects. In order to compare measurements such as these with calculations, the calculations must include defect diffusion processes that enable the defects to migrate to the surface. However, results from the combination of a defect production simulation and a Monte Carlo bulk diffusion simulator, as performed by Jaraiz et al. and described below, could be quantitatively tested with this technique.

C. Defects and dopant diffusion

It was recognized over twenty years ago that the depth profiles of implanted and annealed B could not be explained by either simple range theory or equilibrium diffusion processes. The implant profile broadens by several thousand
angstroms during the early stages of annealing. This phenomenon, known as transient enhanced diffusion (TED), can be seen in Fig. 11 where an initially abrupt profile is subjected to either thermal annealing or shallower implantation damage introduction and annealing. Such anomalously large diffusion rates after implantation were not significant problems when implant depths were in the micrometer range, but they can ruin devices at the shallow implant depths currently in use.

Other anomalous diffusion phenomena have been observed as well, e.g., concentration dependent diffusion and oxidation enhanced diffusion (OED). Understanding these results was achieved by recognizing that dopants in Si must pair with either vacancies or interstitials in order to move; dopants by themselves are immobile. This mechanism successfully explained the concentration dependence of diffusion rate through the charge states of the defects. This mechanism also explains OED and TED by the injection of interstitials during formation of the oxide and from the implantation process, respectively.

Despite a good qualitative understanding of TED and OED, there remain serious problems in quantitatively determining the rates for these processes, largely because the effect depends strongly on the point defect concentrations. Recent experimental results have greatly advanced our understanding of the atomistic mechanisms that govern TED. In the next section, we review these experiments and discuss attempts to develop analytical models of defect and dopant diffusion and to couple defect production simulations with Monte Carlo simulations of defect diffusion. In the final part of this section, we review recent progress in first principles calculations of fundamental defect/dopant interactions that may determine the atomic level mechanisms responsible for the observed behavior.

1. Transient enhanced diffusion

Standard measurements of TED use SIMS to measure the depth profile after annealing (Fig. 4). However, these measurements convolute four phenomena: the implanted damage profile, the stationary component of the dopant profile, the TED itself, and the depth dependence of TED arising from variations in the point defect concentrations causing TED. Over the last two years, much more detailed data on TED has become available using measurements, such as those in Fig. 12, of the TED in superlattice samples. These samples use a regular array of B or Sb doping spikes grown by molecular beam epitaxy (MBE) as marker layers to monitor diffusion. Experiments such as those in Fig. 12 separate implant damage from the diffusing profile, and allow us to quantify the stationary (clustered) part of the profile and the time-averaged diffusivity of a given spike with great precision (because of the extremely abrupt initial spike). By measuring the diffusivity and stationary components of successive spikes, TED and clustering can be quantified as a function of distance from the implant damage in a single...
These measurements reveal the exact time and depth dependence of TED, which relates directly to the effective diffusivity of the point defect profile.

The concentration of Si interstitials as a function of depth from the surface can be directly extracted from the depth dependence of the enhanced B diffusivity shown in Fig. 12. In this way Si interstitial diffusion coefficients can be determined. These data are shown in Fig. 13 along with corresponding OED data. There is good agreement between the experiments. Also plotted are the Si interstitial diffusion coefficients obtained from theory and metal diffusion experiments. These values, while generally consistent amongst themselves, show much higher diffusivities than the data obtained from the B diffusion experiments. Although the diffusivities obtained from these experiments vary by six to eight orders of magnitude, it is now generally accepted that the difference can be explained by the presence of traps. The effective Si diffusivity obtained from the B marker layer experiments is very sensitive to the presence of such traps as C, while the theoretical calculations do not include the presence of traps. It can also be shown that Si interstitial diffusion coefficients extracted from metal diffusion experiments approximate the unretarded values for moderate to low C concentrations. The difference between these two sets of data, therefore, shows the remarkable sensitivity of the diffusing Si interstitial to low levels of traps. A quantitative understanding of the mechanisms of the Si interstitial diffusion and trapping is discussed further at the end of this section.

Along with the improved quantification of the TED process, there is a new push to quantify the state of the extended defect microstructure. Figure 14 shows the impact of quantitative transmission electron microscopy on the study of point defects. Here, TEM has been used to measure the density and size of dislocation loops of known type (interstitial 1/2(110)) during an oxidation step. Dislocation loops were created with an amorphizing implant of 100 keV Si with doses of 2–5 × 10^{15} cm^{-2} as indicated in the figure. As oxidation proceeds, interstitials are injected into the near-surface region, and arrival of interstitials at the dislocation loops leads to loop growth. These measurements provide a near-direct determination of the point defect flux occurring in a given sample region under actual processing conditions, and have been critical in resolving several long-standing problems (such as the circular argument that we know that B is an interstitial diffuser because B diffusion is enhanced by oxidation, and we know that oxidation injects interstitials because it enhances B diffusion). Since the dislocation capture of interstitials is diffusion limited, this technique can be used to measure the defect concentrations independent of a dopant. This is particularly important for scaled device design. At low temperature, dopant diffusion is difficult to use to measure the defect concentrations. However, dislocation loops can best be used as a sensor at these temperatures.

In a related set of experiments, the source of the interstitials driving TED in Fig. 12 was recently identified from quantitative TEM. The detailed diffusion measurements reveal a massive (>1000-fold) enhancement in diffusion in
the initial cascade and subsequent defect annihilation and lower energies and implants at 20–150 keV in the 1

leads to point defect supersaturation and TED. It is also possible to monitor the evaporation of these interstitial condensates during an anneal (see Fig. 16). Such measurements can be correlated with TED data to confirm that the enhancement in diffusion coincides with the evaporation process. This leads to a fairly coherent picture for TED at least for this energy and dose regime where stable interstitial agglomerations form very early in annealing and their evaporation leads to point defect supersaturation and TED. It is also possible to quantify the total interstitial excess associated with these defects in the early stages of the anneal. For Si self-implants at 20–150 keV in the $1 \times 10^{12}-3 \times 10^{14}$ cm$^{-2}$ dose range this interstitial excess appears to correspond to about 1.5 Si per implanted ion, far lower than the Frenkel pair concentrations believed to be associated with the initial implant damage. The measured point defect excess is close to an empirical rule sometimes used in diffusion simulations in which the point defects are assumed to be dominated by the implanted ion (plus one interstitial) rather than the cascade (about 1000 vacancy–interstitial pairs). We are now able to pose a whole new set of questions associated with how the implant damage collapses to this level, what happened to all the vacancies, and what controls the interstitial excess. There are also clear issues relating to how transient diffusion is driven at lower doses (where {311} defects are not observed) and lower energies (where the surface will influence both initial cascade and subsequent defect annihilation).

One of the key new results emerging from the more precise determinations of TED (Fig. 12) is demonstration that the depth dependence of TED can be explained only in terms of trap-limited diffusion of Si self-interstitials. The most probable trap is carbon, an impurity present in all bulk Si to fairly high levels. Identification of a trapping mechanism is perhaps the single most critical outcome of this area. The historical problem with $D_i$ (Fig. 13) can be reinterpreted in terms of variations in carbon content (giving different $D_i$ for very similar experiments) and different approximations for diffusion (the assumption that Si is trap free, giving greatly different results for OED and metal-diffusion measurements of $D_i$) as first suggested by Griffin and Plummer. Moreover, the identification of the trap offers the prospect of not just predicting but controlling TED. Figure 17 compares the TED in B-doped marker layers for MBE growth with and without an acetylene source. Acetylene leads to substitutional incorporation of C at concentrations of $10^{16}$ cm$^{-3}$; at this level, traps not only slow the motion of interstitials, but completely arrest TED. The solid line in Fig. 13 shows the retardation of the Si interstitial diffusivity with a C trap concentration of $1 \times 10^{16}$ cm$^{-3}$.

The connection between this recent data on interstitial sources and historical studies of damage also poses prob-

FIG. 16. Quantitative TEM measurements of evaporation of interstitials from {311} defects (from Ref. 62).

FIG. 17. SIMS measurements of B diffusion profiles in MBE-grown superlattices containing substitutional C levels of (a) $10^{18}$ and (b) $2 \times 10^{19}$ cm$^{-2}$. Profiles are shown for unimplanted and ion-damaged (5 $\times 10^{13}$ cm$^{-2}$; 40 keV Si) superlattices after diffusion in vacuum at 800 °C for 10 min (from Ref. 63).
lems. Traditionally, both amorphization of the target Si and the formation of stable dislocations after annealing have been linked to displacement damage in the cascade. The new picture that is emerging of {311} defects linked to excess interstitials suggests that extended defects may arise from point defect imbalances during the high-temperature portion of the anneal. Evidence is also emerging for a connection between {311} defects and stable dislocations. Recent data show that Frank loops form from a long precursor {311} defect extended in the ⟨110⟩ direction. Such images demonstrate that, at least under some conditions, dislocation formation is also attributable to point defect imbalance. These results offer a new understanding for defect introduction in Si, but also pose critical questions regarding our ability to predict the defect density and spatial distribution in implanted material.

2. Analytical modeling approaches

There have been two primary mechanisms for modeling dopant and defect interactions. Both of these approaches solve a continuity equation for the dopant numerically, but treat the diffusivity of the dopant differently. The first approach is phenomenological and uses an effective time-varying diffusivity that depends on implant and annealing conditions. Expressions for the diffusivity can be derived based on empirical evidence and physical assumptions about defect behavior. The other approach is to solve for the point defect concentrations directly. Additional differential equations are solved that govern the defect behavior and allow the concentration of each defect type to be computed directly. This information is then used to compute the dopant diffusivity.

In theory, the phenomenological approach should be extremely useful since it is simpler to compute and has fewer parameters. While the point defect based approach must characterize all the defect reactions, diffusivities, and time constants, the empirical diffusivity needs only to characterize the dopant motion. The reduced number of parameters should also make it easier to incorporate new effects that occur as junctions are scaled. Several such approaches have been tried and were the first in many cases to be useful for unusual diffusion phenomena. Although it might be expected that this method should be the industry workhorse, in practice this approach is used seldomly, and little research on it is being performed.

Like most empirical techniques, extrapolation beyond the characterized limits of the models is dangerous. Further experimental work must be performed in each new region with additional tuning required for the diffusivity expressions. Although the tuning required can be minimized if the diffusion expressions are rooted in the correct physics, this requires a substantial commitment to the characterization of the junctions in the design space. In one dimension, this is possible since there are adequate measurement techniques for dopant-depth profiling. However, today’s device designers need lateral dopant information to design scaled, submicrometer devices and there are no cheap, widely used methods for two-dimensional profile characterization. Without two-dimensional measurements, there can be no reliable two-dimensional empirical models. This has crippled the development of what would otherwise be the easiest modeling approach. Further research is being conducted on accurate two-dimensional measurement techniques, which could possibly revitalize this approach. A review article by Subrahmanyan covers many of the current approaches, e.g., scanning probe methods, microcontact spreading resistance, sputter methods, and electron microscope techniques.

The other main modeling approach is fully physical and attempts to include all relevant effects. Dopants diffuse through interaction with point defects, and the point defect equations are solved directly. The defect concentrations are then used to compute the local dopant diffusivity. As the defect concentrations change and evolve, so does the dopant diffusivity. Using this system of equations requires a minimum solving three partial-differential equations to account for the motion of one dopant species, and therefore can be computationally complex.

The dopants and defects need to obey the following continuity equation:

\[
\frac{\partial C_A}{\partial t} = \nabla J_A - R_A,
\]

\[
\frac{\partial C_X}{\partial t} = \nabla (J_A + J_X) - R_X,
\]

where \( J \) is the flux, \( C \) is the concentration, subscript \( A \) is the dopant, subscript \( X \) is the defect, interstitial or vacancy, and \( R \) is the recombination. \( R_A \) represents dopant precipitation and may also depend on the defect concentrations. \( R_X \) represents defect recombination of all types. The dopant-defect fluxes are complex and interdependent:

\[
J_A = \sum D_{AX} C_A^* C_X \nabla \log \left( \frac{C_A^*}{C_X^* n_i} \right),
\]

\[
J_X = D_X C_X^* \nabla \left( \frac{C_X^*}{C_X^*} \right),
\]

where \( D \) is the diffusivity, superscript \( * \) represents the substitutional quantity, superscript \( \# \) is the equilibrium concentration, and \( n_i/n_i \) is the electron concentration divided by the carrier concentration. The term \( D_{AX} \) in the first equation is the weighted diffusivity due to the dopant-defect pair. The summation is over both defect types and all defect charge states.

In theory, these systems of equations can predict dopant behavior under a wide variety of conditions. In practice, however, introducing two additional partial-differential equations means that there is a vast increase in the number of parameters. We now require knowledge of the defect diffusivities, defect equilibrium concentrations, recombination rates, boundary conditions, and charge states. A physical model is only as accurate as the physics included, and leaving out any major physical effect can severely limit predictability of the model. As the junctions have scaled, more and more effects and interactions were needed to explain the observed phenomena.

The preceding problems would not represent a major roadblock, except that the defects cannot be measured di-
rectly! This means experiments must be cleverly designed to calibrate the defect models. Interpretation is difficult at best, since the parameters cannot be determined independently. For instance, we most often measure the change in defect concentrations by measuring the change in dopant diffusion, e.g., an increase in the diffusivity of phosphorus is attributed to an enhancement in the concentration of interstitials. However, this same measurement is then used to argue why the diffusivity of phosphorus is enhanced during an interstitial injection. To avoid such circular reasoning, careful experiments must be designed and an equally careful interpretation of the results taken to make progress in parameterization of the defect equations.

3. Monte Carlo simulations of defect kinetics

Atomistic models to study damage accumulation and defect diffusion require modeling over much larger time scales than can be achieved using molecular dynamics simulations. An alternative approach for extending the accessible time scale is to couple the programs for calculation of damage production with a Monte Carlo simulation of the defect diffusion. Recently, Jaraiz et al.\textsuperscript{52} have reported on the development of such a process simulator scheme based on coupling MARLOWE to a diffusion simulator in order to perform simulations of implantation and of high temperature defect annealing for times and temperatures typically used in semiconductor manufacturing.

The simulation is carried out in the following way. A cascade is developed by MARLOWE and the locations of vacancies and interstitials, defined as the set of sites where a displaced atom was in its original and its final position, respectively, are passed on to the MC diffusion simulator. Vacancies and interstitials are then given random jumps, allowing for vacancy–interstitial recombination, annihilation at the surface with a specified sink efficiency, clustering of like defects, and re-emission from the clusters, and trapping of interstitials at native carbon traps. New cascades are added until the specified implant dose is achieved. A high temperature anneal is then simulated by again using the MC diffusion simulator.

A room temperature implant simulation of 40 keV, $5 \times 10^{13}$ cm$^{-2}$ Si into (001) Si, 7° tilt, yields a damage profile consisting of small clusters of interstitials and vacancies reaching a peak concentration of $5 \times 10^{20}$ point defects per cubic centimeter. This value is about five times smaller than that given by MARLOWE alone, since some recombination occurs at room temperature. During the high temperature anneal (815 °C, 5 s), the dominant process is vacancy–interstitial recombination. This continues until all of the vacancies disappear, which occurs in less than 1 s; the vacancies diffuse faster than the interstitials and therefore they reach the surface and are eliminated first. The number of interstitials still present at that time is very small and is close to the number of implanted ions, in agreement with the $+1$ model.\textsuperscript{53} Larger interstitial clusters continue to grow at the expense of smaller clusters which are less stable (smaller binding energies), exhibiting Ostwald ripening as observed experimentally for \{311\} defects.\textsuperscript{62} The contribution of the traps is almost negligible up to this stage. Figure 18 illustrates the interstitial concentration profile for the simulation described at three stages.

In order for these MC diffusion simulations to be accurate and reliable, all the parameters used must be known as accurately as possible. The simulations just described employ migration barriers and cluster binding energies obtained from MD simulations with SW Si.\textsuperscript{35} Ideally, one would like to employ as much data derived from \textit{ab initio} calculations or obtained from experiment as possible. Also, as discussed above, classical MD is better suited than MARLOWE to obtain the primary state of damage and it would be interesting to perform similar atomic level processing simulations with data obtained from MD. Much work remains to be done in this extremely promising area.

4. First principles calculations of fundamental parameters

Progress in the atomistic modeling of dopant and defect diffusion requires a knowledge of the microscopic mechanisms controlling the processes and their fundamental parameters. The coupling of defect and dopant effects in most experiments, as illustrated by TED, together with the complexity of implantation disorder evolution make it nearly impossible to obtain these fundamental parameters by experiment alone.

Recent advances in first-principles theoretical approaches offer new possibilities for progress. For example, improvements in \textit{ab initio} calculations based on the local density approximation, in combination with the rapidly increasing computational power of workstations and massively parallel machines, allow Si defect and dopant calculations of increasingly realistic size and accuracy. In this way theorists can work hand in hand with experiment and TCAD modeling to help interpret results, evaluate proposed mechanisms, and establish parameter values. Key measures of success are the reduction in the number of empirical parameters and the extraction of relevant physically based parameters.
In a hierarchical scheme of applying theory to ion implantation the first-principles methods offer the possibility of calculating stable configurations, energy barriers, and minimum energy paths. Even when absolute values cannot be given with certainty, the ability to obtain good relative values will be very useful in combination with experiments and models in sorting out the possible mechanisms and developing correct physical descriptions. Dopant and defect interactions can thus be evaluated as they determine diffusion, clustering, recombination, dissolution, and related events, offering the possibility of defining much more accurate rules for the MC descriptions. Such treatments will be critical to an eventual fully predictive modeling of the transient enhanced diffusion process.

First-principles studies along the above lines for Si are now beginning in a number of laboratories. An example for the case of the B-interstitial interaction is illustrated in Fig. 19.51,84 In the case of boron transient enhanced diffusion occurs predominately by way of Si interstitials and the “kick-out” mechanism is believed to be the dominant mechanism. This process involves a Si interstitial moving onto a Si lattice site and ejecting a substitutional boron atom which then migrates rapidly for some distance before being recaptured onto a Si lattice site and releasing a new Si interstitial. Ab initio calculations provide a means to evaluate the migration paths and barriers involved in the process. These values then allow rate equation evaluations for comparison to the observed diffusion lengths, anomalou temperature dependences, and other properties, thereby lending credibility to the proposed kickout mechanism in this case. Other ab initio calculations are evaluating vacancy interactions with group V dopant interactions with vacancies, and the systematic binding energy variation from P to As to Sb.85

It is important to note that even the seemingly simpler self-diffusion problem in Si is not yet resolved. Comparison of the defect formation and migration energies with experiment is simplified in this case and experimental data have been published for self-diffusion in Si measured by radioactive isotopes under equilibrium conditions. The measurements exhibit Arrhenius behavior over a wide temperature range, with an apparent activation energy of 4.8±0.2 eV and a very large prefactor of 10² cm²/s.40 If the self-diffusion jumps are induced by interstitials, then the coefficient of self-diffusion is \(D_i C_i\), where \(C_i\) is the equilibrium interstitial concentration normalized by the Si atomic concentration. The apparent activation energy for this mechanism is \(E_{I,F} + E_{I,M}\). Note that the classical MD interstitial calculations give \(E_{I,F} + E_{I,M} = 4.5\) eV, almost within the range of error of the experiment. But the fastest mechanism will predominate, and MD calculations suggest vacancies should prevail at all temperatures, with \(E_{v,F} + E_{v,R} = 3.0\) eV. This energy is far below the experimental values. (The entropy of formation for vacancies and interstitials is of the order of unity,86 and thus would not have a strong effect on the equilibrium concentrations; therefore, the formation energies and diffusion coefficients should provide a good estimate of the relative contributions of the two mechanisms.) Clearly, much research remains to be done in this area in order to bring the computational and experimentally measured values of the Si defect properties into agreement. To fully exploit the first-principles calculations and develop extensions of the atomic level simulators for treating dopant/defect interactions, dopant clustering, and precipitation extensive new input data must be obtained from experiments for validation of state-of-the-art ab initio calculations and correlation with parameter modeling.

III. DEEP IMPLANTATION: BURIED STRUCTURES AND GETTERING

While much of the technological and scientific emphasis of implantation focuses around the use of low energies for shallow junction formation, there continues to be a burgeoning interest in the use of higher energy beams in at least three disparate areas. The area that industry has adopted most readily has been the use of high energy beams to directly fabricate retrograde wells or tubs. These deep dopant implants obviate the use of lengthy high temperature anneals to drive the dopants in by diffusion. Moreover, a significant number of processing steps can be eliminated with concomitant cost saving and yield improvement. This usage is leading to new processing protocols where clustered, low and high energy implants are performed prior to an annealing or gate oxide growth step. Understanding the interaction among these clustered implants will become increasingly important. We discuss some of the defect issues below with particular emphasis on the extended defects that can form in high energy implantations and can thread to the surface.

The production of buried dielectric layers by high dose, high energy implantation into Si has been studied for some twenty years now.87 The goal is to produce a dielectrically isolated Si layer with sufficient quality for device fabrication. Most work has focused on the use of oxygen implants and the direct production of a buried oxide layer by separation by implantation of oxygen (SIMOX). Much progress has been made in recent years in reducing the defect density by tun-

FIG. 19. The results of first principles calculations for the energy barriers associated with B TED. (a) Schematic of a section of the (110) plane in Si illustrating the B replacement and migration associated with this process. (b) Calculated energy barriers for the Si interstitial kick-out mechanism (after Refs. 41 and 84).
ing the implantation and annealing steps so that large scale integrated circuits that run at increased speed and lower operating voltages can now be fabricated. In the following sections, we will review some of the defect formation and agglomeration mechanisms in the overlying Si after the direct formation of the buried layer by O implantation. Unlike the low energy regime, vacancy agglomeration assumes an important role in defect formation for implant processes.

In addition to the SIMOX process, alternative methods for producing buried dielectric isolation layers have advanced rapidly. In the process of wafer bonding, two wafers are joined by placing their oxidized surfaces in contact with each other and annealing them. The resulting layered structure is then thinned from one side by chemical-mechanical polishing and etching to form a thin Si layer above the buried oxide. More recently, a technique was introduced using a combination of wafer bonding and ion implantation; this technique holds much promise for device fabrication. In this Smart-Cut technique, one of the Si wafers is initially implanted with a high dose of H before the surface oxide is formed. The wafers are then weakly bonded together at room temperature. The bonded structure is subsequently annealed in a two-step process. The first anneal at 400–600 °C causes the implanted wafer to split apart into a thin layer of Si (bonded through the oxide to the second wafer) and the remainder of the implanted wafer. A high temperature anneal strengthens the bonding of the thin Si layer to the silicon-on-insulator (SOI) structure. Because the splitting occurs at the peak of the depth distribution of the implanted gas, the thickness of the Si overlayer can be controlled easily by the energy of the implant. The final area we will review in this section is gettering. Gettering describes the phenomenon whereby deleterious impurities, such as Fe, can be removed from the active area of the Si wafer and be trapped in treated regions. Ion implantation is proving a particularly interesting way of treating certain areas to introduce well controlled gettering sites. We will discuss gettering at nanocavities, implanted impurities, and ion-induced damage. Although there is still debate about the commercial viability of implantation gettering for device fabrication, there is little doubt that the ability to accurately control and measure the gettering profiles through implantation and SIMS analysis is leading to a better understanding of these complex phenomena.

A. Well formation

1. Processing constraints

The introduction of high energy implantation into production is being driven by well- or tub-formation technologies. CMOS manufacturing requires both n-channel and p-channel transistors, and the process begins with formation of the n- and p-type wells (or tubs), several micrometers deep. Traditionally, this is accomplished by thermal diffusion drive-in anneals at high temperature (up to 1100 °C) and over prolonged times (many hours). The use of high energy implants for well formation offers several important advantages. Most notably, the high energy implantation well process can eliminate two or three mask steps from the total process sequence (see Fig. 20). This process simplification is essentially a consequence of our ability to control the depth of the implanted dopant, and implant it through a pre-existing field–oxide isolation layer. For triple-well processes in low-voltage 16 MB dynamic random access memory (DRAM) replacement of a six-mask diffused well process by a four-mask high energy implantation process has been reported. The mask steps refers only to the photolithography process steps (the most expensive), but a number of associated steps is also eliminated, for the further cost reduction. For example, total cost savings of approximately 10% have been estimated for 16 MB DRAM manufacture. This 10% cost reduction can amount to around $70 per wafer for 200 mm wafers: with a throughput of about 3000 wafers per week a high energy implantation step can achieve a total cost savings amounting to up to $0.2 M/week in a large fabrication facility.

In addition, this reduced process complexity can also increase yield in the production line and shorten the cycle time. In general, yield models assume a fixed probability of failure (or fixed particle contamination probability) at each process step. Smaller numbers of steps then imply smaller numbers of particles or higher yield through the line. The shorter cycle times can also be important drivers for introduction into manufacturing. Thus process simplification alone provides good reasons for replacing diffused tubs with high energy implantation. For the increasingly large wafer diameters in use, the reduction in thermal budget alone is also significant. For larger diameter wafers, warpage can be
an issue at the high temperatures as can long times involved in tub diffusion: high energy implantation reduces both time and temperature by putting the implant at the depth of the desired dopant distribution. This advantage becomes more significant as wafer diameters grow.

Finally, high energy implantation offers considerable advantages in terms of device performance. From a device point of view, high energy implantation technology allows precise positioning and control of dopants and junctions with minimal vertical and lateral diffusion. Reported improvements in device characteristics include: reduced junction leakage, lower soft error rates, latch-up immunity, increased DRAM retention time, and reduction in single bit failure. At the circuit level, improved low voltage operation, operating speed, flash chip size reductions (30%–50%), and overall increase in circuit yields have also been reported.

2. Extended defect formation

With the advent of the extensive use of high energy implantation, extended defect formation from the implant is becoming an issue for the first time in years. For low energy implants dislocations are not generally a problem; either the dose is so low that there are no extended defects after the activation anneal (source–drain extensions, threshold adjust implants), or the electrically active junction lies well below the defective region (source–drain implants). (Note that TED is almost beneficial here, since without anomalous diffusion the source–drain junctions would lie perilously close to the dislocations from the end-of-range of the source–drain implant.) With high energy implantation, we are moving into a regime where the dose must be far higher than a threshold adjust implant, but the implant lies below the active region of the device. Dislocations from the implant then have the possibility of extending towards the surface. A single dislocation penetrating the depletion region of any p-n junction anywhere on a 1 cm² chip will cause a leakage failure. Thus, in manufacturing we are unable to tolerate a density of such dislocations higher than about 0.1 cm⁻². Understanding of dislocation formation processes is insufficiently complete to allow us to predict with any confidence the precise implant and annealing conditions needed to meet this requirement.

Although a number of studies (e.g., Ref. 69) has investigated the “threshold dose” for dislocation introduction, there are very few studies providing quantitative data, elucidating mechanisms, or studying the more relevant near-surface class of threading dislocations. Very recently, a quantitative picture has begun to emerge. Figure 21 (taken from Ref. 93) shows the typical dislocation microstructure for MeV implants following a high temperature anneal. As is usual with higher energy implantation, these dislocations are of the perfect a/2(110) Burgers vector type. The appearance of very long, extended dipoles represents a higher energy microstructure than elliptical loops. There is, as yet, no clear consensus on the mechanism for formation of these defects, although it is known that perfect dislocations are associated with high energy implants while very low energy amorphizing implants more frequently show 1/3(111) Frank loops. There is reason to believe that dislocation formation is associated with the same {311} defects that cause TED (Sec. II). The dislocation microstructure seen in Fig. 21 does indeed mirror very closely the appearance of {311} rodlike defects extended along (110) directions. This is significant because it suggests a link between the dislocations and interstitial supersaturations, as opposed to displacement damage. As for low energy implants, our understanding of the point defect supersaturations remains incomplete (high energy implants may have the additional complication of a spatial separation between interstitial and vacancy excesses, as will be discussed in Sec. III C).

The length of a typical dislocation dipole in Fig. 21 clearly exceeds the depth of the implant (2 µm for the implant in Fig. 21), so there is obviously a strong possibility of threading dislocations penetrating all the way to the surface. Defect etching studies of threading densities show extremely high dislocation densities (~10⁶ cm⁻²) at doses of about 1×10¹⁶ cm⁻². Either lower or higher doses reduce this density, as illustrated in Fig. 22. The falloff at low dose is associated with the threshold for dislocation formation. At high doses the reduced threading dislocation density is associated with an increase in the total number of dislocations,
but a modified dislocation microstructure that leads to much shorter dipoles that no longer extend to the near-surface region. The detailed physics of all these interactions is not yet understood; this remains one of the critical challenges for our understanding of high energy implantation. Figure 22 illustrates the importance of understanding different mechanisms: by preannealing, it is possible to reduce the dislocation density even at \(1 \times 10^{14}\) cm\(^{-2}\) implantation doses; in this instance the reduction seems to be attributed to the O precipitation accomplished by a preanneal.

Another example of the resulting complexity that can occur through impurity–defect interactions is given in Fig. 23 for oxygen accumulation in high energy implant damage.\(^{94}\) At doses of \(1 \times 10^{14}\) cm\(^{-2}\) and above, oxygen pileup at these residual damage sites has been reported; for a dose of \(>1 \times 10^{15}\) cm\(^{-2}\), two oxygen peaks are observed as shown in Fig. 23(a). This second peak at about one-half the projected range (\(R_p\)) is speculated to be vacancy related, and the heavier the implanted ion mass the larger the oxygen peak, as shown in Fig. 23(b). However, if a denuded zone, free of oxygen, is formed in the wafer prior to the high dose implantation, then oxygen pileup does not occur.

B. Defect issues in buried oxide formation

We now review results obtained from the formation of buried oxide layers by the SIMOX process and the increased understanding of high energy defect formation that has resulted. Attention is focused upon the overlying Si layer (above of the buried oxide) where the behavior of damage growth reveals interesting aspects of the defect interactions not seen (or fully appreciated) during irradiation under less severe conditions (i.e., at lower energy, dose, or temperature). Unlike the TED phenomena, this behavior is dominated by vacancy defects. While this vacancy-dominated behavior may be more readily apparent under the extreme conditions inherent to SIMOX processing, experimental evidence suggests that it is present over a much wider range of implantation conditions.

SIMOX is a process for forming a high-quality, SOI material\(^{88,95}\) by implanting a single crystal of Si with O ions to sufficiently high doses to produce a buried oxide layer. Implantation is generally done at elevated temperatures to prevent amorphization of the surface Si layer and to promote dynamic annealing of ion-induced defects. Advantages of SOI over bulk Si as a substrate material for use in fabricating

FIG. 23. (a) Oxygen depth profile after high energy implant. Note pileup of oxygen at half the projected range. (b) Relation between oxygen concentration gettered by implant damage at half the projected range and ion mass (from Ref. 94).

FIG. 24. RBS spectra of the aligned yield along the normal axis from Si(100) implanted at 450 °C with 500 keV O ions at two different fluences. The spectra shown by the open and closed circles were implanted continuously in a “single step.” The “two-step” process involved interrupting the O implantation as described in the text. All O implantations were done using a current density of \(5–6\) mA cm\(^{-2}\) which raised the temperature of the sample during irradiation to \(\sim 530\) °C. An aligned spectrum from virgin Si and a random spectrum from an implanted sample are shown for reference.
integrated circuits include improved device and circuit performance and reduced wafer processing cost. The unique behavior of the damage produced during SIMOX-type implantations is demonstrated in the spectra in Fig. 24 from Si implanted at 450 °C with 500 keV O ions at two different doses (designated as one-step implants). The overlying Si layer for these samples extends to ~0.6 μm in depth with the buried oxide layer formed beyond, near the ions’ projected range at ~0.9 μm. At the lower dose, the aligned yield within the Si overlayer remains near the virgin level, indicating little or no accumulation of damage. The absence of damage accumulation at this dose is surprising since the average number of atomic displacements (dpa’s) within this region is nearly 100 (as determined by TRIM) and indicates that an overwhelming majority of the Frenkel defects recombines and does not contribute to damage growth. At the higher dose, the damage behavior within the overlayer undergoes a transition from constrained to rapid growth, as indicated by the substantial increase in the aligned yield within the top ~0.5 μm. Cross-sectional transmission electron micrographs (XTEMs) from these two implanted samples are shown in Figs. 25(a) and 25(b) at the lower and higher doses, respectively. The morphology near the projected range is not unusual and has been described previously. The features within the Si overlayer provide clear evidence of a transition in the damage behavior. At the lower dose, defects are too small to be resolved, whereas after the transition at the higher dose, the defects consist of a band of dislocations near the middle of the overlayer which is isolated from both the surface and the end-of-range defects. The critical dose marking the onset of the transition was found to increase with temperature (e.g., the transition was observed near 7×10^17 cm^-2 for 450 keV O implantation at 250 °C, vs 1×10^18 cm^-2 at 450 °C in Fig. 25). Dislocations formed at the onset of the transition first appear as loops but can quickly evolve into dislocation tangles (especially at lower temperature). The dislocation density within these tangles decreases with temperature such that above ~600 °C they remain as isolated (half) loops.

The behavior of precursor defects formed within the overlayer is key to understanding the mechanism that triggers the onset of the transition. In order to observe these defects at low doses where XTEM is unable to resolve the individual defects, the lower dose sample was annealed at 900 °C to enlarge any defect clusters present due to ripening. The micrograph in Fig. 25(c) reveals a band of microvoids extending over a range similar to that of the dislocation band formed by the transition [as seen in Fig. 25(b)]. These voids clearly evolve from small vacancy clusters that must be present within the region at fairly high concentrations (i.e.,
well above the equilibrium value). X-ray diffraction analysis demonstrated that these small vacancy clusters produce a uniaxial contraction of the lattice along the sample normal which increases with implantation dose up to the damage transition, after which there is some strain relief. Therefore, it appears that a buildup of vacancy clusters, as well as the concomitant strain, continues during irradiation until a critical value that triggers the damage transition is achieved. This model is supported by comparing the results of the one-step process and the two-step process in Fig. 24. In the single-step process (discussed above), O implantation was done continuously until the desired dose was achieved. In the two-step process, O implantation was interrupted after an initial subcritical O dose (below the damage transition) to irradiate the sample with Si self-ions at an energy (150 keV) that placed them near the middle of the superficial layer where the density of vacancy-type defects is greatest. The introduction of these self-ions acts as an artificial recombination tool to eliminate a portion of the vacancy population, and thus the associated strain. Last, a second step of O implantation was used to achieve the final dose. Comparison between these processes shows that, for a total dose of $1.1 \times 10^{18}$ cm$^{-2}$, the damage transition within the Si overlayer (seen in the one-step spectrum) did not occur during two-step processing. This demonstrates that the transition can be delayed or avoided by direct manipulation of the vacancy defects by self-ion irradiation which, in turn, establishes the vacancy-type defect as the precursor that triggers the damage transition.

These results were compared with damage effects using self-ions under similar conditions to determine whether or not vacancy mediated damage growth is derived from impurity effects (due to high concentration of oxygen) or is of a more general origin. The results of implantation at 450 °C with 1.25 MeV self-ions are shown in Fig. 26. Aligned spectra from samples implanted at two different doses display similar features to those in Fig. 24. The lower dose ($2.0 \times 10^{17}$ cm$^{-2}$) self-implanted sample shows two distinct behaviors of damage growth; near the end-of-range where substantial growth occurs and an overlayer region in which growth is constrained. The constrained growth within the overlayer persists to quite high doses but, as in the case for O implantation, it eventually yields at higher doses. At a self-implanted dose of $3 \times 10^{17}$ cm$^{-2}$, a distinct peak near the center of the superficial layer at $\sim 0.45 \mu$m signifies significant growth of damage. The similarity between these results and the O implantation confirms that the vacancy mediated growth of damage is not simply a consequence of the ion-solid chemistry.

One explanation for the observed excess of vacancy-type defects within the overlayer prior to dislocation formation is the spatial separation of Frenkel defects. Since interstitial–vacancy pairs are created during irradiation as a result of atomic displacements, the momentum transferred to the interstitial will on average have a nonzero component along the direction of the incident ion. At these higher energies, this leads to a spatial separation between the interstitial and vacancy defects with the interstitials distributed to greater depths on average. TRIM simulations which track the trajectory of the primary ions as well as the knock-ons have been used to calculate the separation in the distributions of the Frenkel pairs. The results, shown in Fig. 27 normalized to a dose of $10^{18}$ cm$^{-2}$, reveal a vacancy excess of several atomic percent over the top 0.8 μm followed by a much narrower distribution of excess interstitials. The excess concentrations are converted to volumetric changes within the solid by associating an atomic volume with each of the pairs (i.e., a missing lattice site for a vacancy and an extra atom for an interstitial). Since the excess derives from small differences in the Frenkel pair distributions, it is nominally only a fraction (~0.03%) of the total defect population. However, the majority of the Frenkel pairs is spatially correlated and rapidly recombines at elevated temperature. In contrast, the spatially separated pairs (i.e., the excess defects) avoid recombination and, therefore, contribute to a very slow rate of damage accumulation. Also shown in Fig. 27 is the volume added by the implanted oxygen assuming it precipitates as an
oxide, a reasonable assumption since most of the $10^{18}$ cm$^{-2}$ implanted oxygen exceeds the solubility limit in Si of $\sim 10^{17}$ cm$^{-3}$ at 600 °C. The sum of these two effects yields a crossover depth between rarefaction and densification at 0.55 $\mu$m as indicated by the arrow in Fig. 27. This thickness coincides well with the extent of the vacancy-rich overlayer region shown in Fig. 24. While TRIM is an athermal calculation and does not include effects of defect diffusion, substantial interdiffusion of the interstitial and vacancy excesses is not anticipated. The results clearly show that these defects are self-trapping and thus produce immobile clusters. Also, estimates of the vacancy cluster size based on the size dependent surrounding lattice distortion and measured layer strains give cluster sizes on the order of six vacancies, consistent with expectations for clusters that are too small to resolve by TEM and yet sufficiently large to be stable at $\geq 450$ °C.

Since the pair separation mechanism derives from the kinematics of the ion–solid interaction, it is inherently independent of temperature. However, at lower temperatures the pair separation mechanism may be obscured by other damage mechanisms and thermal kinetics will determine the nature of subsequent interactions among the excess vacancies produced by this mechanism. The behavior of the ion-induced damage growth at room temperature (RT) is shown by aligned spectra in Fig. 28 from Si(100) crystals implanted at higher ion energies (1.25 MeV) with different doses of self-ions.$^{108}$ Two distinct behaviors of the damage growth are immediately obvious. Of particular interest is the behavior within the overlayer region ahead of the ion end-of-range ($\sim 1.3 \mu$m). The aligned yield from this region does not change over the range of dose shown but remains at a low level, indicating that damage growth saturates at a low concentration within this region during irradiation. This contrasts with the behavior near the end-of-range where damage is shown to increase monotonically with dose. A more detailed description of the damage behavior at RT can be found in Refs. 108 and 109. This dichotomy in the behavior of damage growth is very similar to that previously described for much higher temperatures and suggests, at the least, that there are some common mechanisms that span the range of temperature.

### C. Impurity gettering

For successful fabrication of device structures, the near-surface region after processing must be free of oxygen, oxygen-related defects, as-grown crystal surface imperfections, and transition metal impurities. If such a defect-free zone is not properly incorporated into the design and formed, oxygen precipitates and crystallographic defects may be located in the device is active region and result in device degradation. Transition metal impurities are known to cause detrimental effects when present in the active regions of Si devices,$^{110–112}$ thereby degrading minority carrier lifetime, providing recombination-generation centers, increasing junction leakage current, and reducing gate oxide integrity. A potentially important use of high energy implantation is for the controlled introduction of dopants, impurities, and damage for the implementation and study of gettering.

In current technology, two gettering schemes are used: intrinsic gettering and gettering by epitaxial layers. Intrinsic gettering involves the processing of the Si wafer in such a way as to make available internal gettering sites (such as heat treating to form oxide precipitates). It has been optimized through widespread use and is presently a standard manufacturing process. This scheme involves temperature cycling Czochralski (CZ) Si to create oxide precipitates in the bulk of the wafer along with a defect free, low impurity denuded zone near the surface for the device active area. Intrinsic gettering works by providing nucleation sites at strained internal interfaces around the internal oxide precipitates which cause the precipitation of the metal impurities (typically as metal silicides). Often, this approach incorporates existing high temperature processing steps to optimize the formation of the denuded zone and the bulk oxide precipitates. The other widely used gettering scheme is that of epitaxial wafers (e.g., $p$ on $p^+$ substrates). These wafers were originally used for latch-up suppression but were demonstrated to have high gettering efficiencies. This high gettering efficiency is believed to occur through the action of both the B and O in the highly doped substrate.

For future device generations, integrated circuit processing will necessitate much lower thermal budgets to achieve the required reductions in feature sizes. At the same time the acceptable level of metal impurities will steadily decrease, (see Table I),$^{12}$ so that for some next generation circuits metallic impurity limits are expected to be approaching $10^9$ atoms cm$^{-2}$. These combined trends raise special concerns for intrinsic gettering. Not only will the formation of the precipitated oxide and denuded zone be difficult to achieve with the lower thermal budgets, but the solid solubility of transition metals may make gettering by silicide formation inadequate to reach the needed concentration levels. As a result, extrinsic gettering techniques are currently being explored as alternatives for cleaning up the near-surface region. In extrinsic schemes, the gettering sites are
externally introduced, e.g., by introducing damage or phosphorous diffusion on the backside of the wafer or by frontside patterned or buried implantation treatments which would be more effective at lower thermal budgets. Figure 29 shows a cross-sectional schematic of a denuded zone wafer with various implant gettering schemes.

Ion implantation damage was shown to getter metal impurities more than two decades ago. Although various groups have explored implantation as an extrinsic gettering technique, such studies have not been exhaustive. In light of current processing trends, several inherent advantages of ion implantation have led to interest in re-examining this approach. Ion implantation is a standard process tool for doping and is inherently compatible with frontside (proximity) methods for introducing gettering sites. Further, the advent of high energy implantation provides the possibility of introducing damage or impurity species into the Si below the active device region (see Fig. 29). It can also readily be used in conjunction with lateral patterning for the selective introduction of gettering regions. In the following section, we discuss three new gettering approaches based on ion implantation: (1) the deeper introduction of extended defects (such as dislocation loops), (2) the introduction of three-dimensional cavity defects with new gettering characteristics, and (3) the introduction of impurities (such as carbon) with stronger gettering behavior.

1. Implantation damage

The gettering of metal impurities at ion implantation produced damage in Si was first demonstrated in the early 1970s. Impurities such as Fe, Co, Ni, Cu, and Au were shown to be getterted by a variety of implanted species. More recently, channeled and high energy Si self-implantation has been examined in more detail for its gettering properties. End-of-range dislocation loops and other defects provide gettering sites, e.g., regrown buried amorphous layers formed by Si implantation leave a layer of defects where the two internal crystalline–amorphous interfaces meet upon solid-phase epitaxial regrowth. These defects have been demonstrated to effectively getter Cu from the Si surface region with the process being most efficient at 600 °C. Other studies have focused on proximity gettering of Fe and have demonstrated through reduced junction leakage currents the electrical benefits of high energy Si self-implantation for gettering heavy metals. In general it is found to be more difficult to getter Fe than Cu impurities. Also, use of boron implantation, as is standardly employed for high energy well implants, is not found to provide for effective intrinsic damage gettering, presumably due to the lesser defect introduction.

An inherent advantage of purely intrinsic damage gettering due to Si implantation is the lack of the introduction of any other impurity or dopant species. However, the gettering mechanism is not fully understood and it is not yet apparent that sufficiently effective gettering can be achieved in this manner for future generation circuits.

2. Implantation-formed nanocavities

A recently developed approach to gettering is the formation of a layer of microscopic cavities within the Si (Ref. 120) by high fluence He implantation. Upon annealing to 700 °C or higher, the He outgases from the Si, leaving behind microscopic cavities (see Fig. 30). These cavities are highly stable and persist to processing temperatures as high as 1200 °C. The internal free Si surfaces of these cavities are found to provide strong traps for transition metal impurities and offer the possibility of a new gettering mechanism by chemisorption onto the internal cavity surface. Since it is based on a trapping mechanism, rather than second phase formation, this approach offers the possibility of gettering impurities to much lower concentrations than the metal silicide precipitation used in intrinsic gettering. Gettering by helium implantation induced nanocavities has been shown for Cu, Ni, Co, Fe, and Au, and this process has now been demonstrated by a number of groups.

The formation of gas bubbles due to inert gas implantation has been known for some time, but the role of these open volume defects in gettering in Si was only recently identified. From TEM observations the ripening of the cavities upon annealing was inferred to be due to internal migration and agglomeration of the cavities rather than to dissolution of vacancies. Extensive studies of cavity proper-
ties have been carried out, including the effects of internal faceting and of their electrical and hydrogen trapping properties.

The trapping of Cu at nanocavities has been studied in detail and had led to a quantitative understanding of this gettering process. In Fig. 31 an example is shown of such studies, where the Cu was introduced by implantation into the backside of a Si wafer and He implantation used to form nanocavities in the frontside of the wafer. Heat treatment results in Cu silicide formation and subsequent heating leads to the transport of Cu across the wafer and trapping within the nanocavities. The results of trapping versus time have been quantitatively modeled using the diffusivity and solubility of Cu in Si and the heat of solution of 1.7 eV for Cu relative to Cu$_3$Si. The saturation level of Cu within the nanocavities is found to agree well with the expected number of bonding sites available at the cavity surface as determined by TEM measured cavity surface area and assuming seven Si bonds per nm$^2$ available on $\{111\}$ and on $\{100\}$ Si surfaces. These results provide strong support for the assumption of chemisorption bonding of Cu on the cavity walls. Numerous experiments for various temperatures and layer combinations have been used to confirm this interpretation.

The process of chemisorption of the metal impurities onto the internal cavity surface has important consequences in terms of the gettering process. Trapping due to chemisorption offers a thermodynamic advantage over precipitation in that trapping will persist to arbitrarily low concentrations, maintaining a constant ratio between the amount of impurity remaining in the solution and the fractional occupation of the trapped sites. Based on this mechanism an example is shown in Fig. 32 that demonstrates the predicted enhancement of many orders of magnitude for the gettering of Cu by the nanocavity mechanism compared to conventional silicide precipitation gettering at low temperatures.

Studies have been performed for Cu, Ni, Co, Fe, and Au for nanocavity trapping. Gold is found to be trapped up to approximately one monolayer of internal cavity surface, similar to Cu. However, Ni, Co, and Fe are found to trap to lower levels due to stronger binding in the silicide phase relative to the chemisorption binding energy (see Table II). Even under these conditions however, there is still a sizable binding energy ($\sim 1.5$ eV) for the nanocavities and

![FIG. 31. (a) Schematic showing implantation of Cu into the backside of the wafer and He implantation into the frontside to form nanocavities. (b) Diffusion of the Cu from the implanted silicide to the cavities enables determination of the relative energies for silicide formation and trapping at internal cavity surfaces.](image)

![FIG. 32. Relation between Cu concentration after gettering and before gettering. Gettering of Cu by nanocavities is predicted to be many orders of magnitude greater than conventional silicide precipitation at low temperatures.](image)

<table>
<thead>
<tr>
<th>Metal impurity</th>
<th>Nanocavity ($\Delta G$)</th>
<th>Silicide ($\Delta G$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu</td>
<td>2.2</td>
<td>1.4</td>
</tr>
<tr>
<td>Au</td>
<td>2.6</td>
<td>2.3</td>
</tr>
<tr>
<td>Co</td>
<td>1.6</td>
<td>2.2</td>
</tr>
<tr>
<td>Fe</td>
<td>1.5</td>
<td>2.3</td>
</tr>
</tbody>
</table>

TABLE II. Summary of nanocavity gettering measurements for various impurity metals at 700 °C in Si. The results given are the impurity binding free energies, $\Delta G = \Delta H - T\Delta S_{sp}$, for cavity surface chemisorption expressed relative to solution in the Si lattice (data from Refs. 12, 122, 127, 128).
thus substantial fractions of metal impurities such as Fe can be gettered relative to levels that would be expected to be present for Si wafers under processing conditions. These results suggest that this new gettering mechanism may provide more effective trapping than that available by conventional approaches, and may, as well, extend down to much lower concentration levels than previously available. The results also illustrate the possibilities of developing more quantitative descriptions of gettering processes whether using implantation or other approaches.

Several other relevant features have been established for nanocavities and they would be important considerations in an actual processing application. First, the entities also serve as nucleation sites for silicides at higher levels of impurities. In addition, these internal surfaces strongly bind hydrogen, and in fact the surface Si–hydrogen bond energy (2.6 eV) has been established through cavity experiments.121 This strength is sufficiently large to displace the bonding of Cu at cavity walls in the presence of hydrogen, and the process is reversible (i.e., if the H is removed, the Cu can be retrapped at a nanocavity surface). It also has been shown that for CZ Si wafers, with their higher O content, some O can be trapped by the cavities, although a substantial number of sites remained available for metal impurity trapping.122 Finally, it should be noted that the electrical properties of cavity surfaces in Si have been investigated and their deep levels established (they can charge either positively or negatively by trapping electrons or holes depending on the Fermi level position).121 Thus, electric fields within several Debye screening distances of nanocavities may be present depending on the particular conditions experienced and they need to be taken into account in the case of short distance proximity gettering.

3. Implanted impurities

Another alternative gettering technique is MeV implantation of various impurity atoms such as oxygen and carbon at high doses of $1 \sim 5 \times 10^{15}$ cm$^{-2}$.117 A cross-sectional model of a silicon wafer with high energy implants and low temperature gettering is shown schematically in Fig. 29. A denuded zone 5 μm deep depleted of oxygen is shown with bulk intrinsic gettering and backside extrinsic gettering. The MeV implanted gettering layer is below all of the high energy well implants and end-of-range damage and 1/2 $R_p$ point defects are also indicated. This buried gettering layer can also be produced with phosphorus and boron where rapid interstitial oxygen (O$_i$) diffusion and pileup at residual implant damage have been observed to create a denuded zone 1–4 μm deep.129 A recent study has used high energy B implantation into Si wafers contaminated with known levels of Fe to quantitatively elucidate the gettering mechanism.130 The depth profile of Fe throughout the wafer was measured using both deep level transient spectroscopy (DLTS) and SIMS in the high concentration region of the B implant. Figure 33(a) shows DLTS signals from the front- and back-sides of the wafer as a function of B dose. The B implant at a depth of 4 μm very efficiently removed the Fe in the near-surface region (the DLTS probes the first 2 μm of the Si) but did not influence the Fe concentration in the bulk of the wafer. Figure 33(b) shows the Fe which is gettered at the implanted B. This study unravels the kinetics and thermodynamics of the Fe–B gettering process. The gettering action occurs during cooling through the Fermi-level-enhanced pairing reaction between Fe and B which is operative at temperatures lower than 500°C.

Carbon implantation into Si was recently reported to be an effective impurity species for Au, Fe, and Cu gettering.117,125,131 In one study surprisingly low fluences of $10^{12}$ C cm$^{-2}$ were reported to be effective for Fe gettering as measured by p-n junction leakage currents for high energy (1.3 MeV) carbon implants,131 but this remains to be confirmed. It was also shown that the C implantation is much more effective than oxygen and that, given sufficient He doses to form nanocavities, these latter traps are significantly more effective than the C or O for gettering.125 The C was inferred as not to be simply enhancing the nucleation of oxide precipitates for its effect. Carbon is also known to have a number of interesting effects on implantation damage, in-

---

FIG. 33. High energy B implantation effectively getters Fe at depth of 4 μm. (a) DLTS measurements from front- and backsides of the wafer as a function of B dose. (b) SIMS depth profile of Fe gettered by implanted B (from Ref. 130).
cluding promoting the reduction of end-of-range defects. The influence of carbon on gettering is thus of considerable interest, although the microscopic nature of the gettering process is not understood at this time. The major potential advantage of impurity gettering by species like C is the ability to achieve gettering at lower fluences than may be required for nanocavities. However, the carbon induced gettering centers must have sufficient stability if they are to be used in semiconductor processing. This knowledge, as well as a quantitative understanding of the process (e.g., binding energies and kinetics) still remains to be determined for the potential of this approach to be established.

IV. ION BEAM ANALYSIS: SECONDARY ION MASS SPECTROMETRY AND RUTHERFORD BACKSCATTERING SPECTROMETRY

Progress of ion beam processing techniques, such as implantation, has been closely linked over the years with the development of ion beam analytical techniques, such as Rutherford backscattering spectroscopy (RBS) and SIMS. Progress in both areas is linked by the common need for precise control and reproducibility of beam parameters and quantitative understanding of the physical mechanisms of ion–solid interactions. The early demands to measure dopant depth profiles in Si played a considerable role in the development of SIMS and RBS. SIMS analysis can now routinely characterize one-dimensional depth profiles in the near-surface region (2–100 nm), while backscattering spectrometry analysis at high energy (>3 MeV) or medium energy (~100 keV) allows detection of high mass elements and surface contaminants.

The need for characterization of shallow-junction dopant profiles in integrated circuit technology is, however, once again creating severe challenges for these analytical techniques. Neither of the two classical techniques has been applied successfully to determine two-dimensional profiles typified by the lateral spreading of dopants under the gate regions of field effect transistors. In this section, we review the current status of these ion-based analytical techniques with emphasis on their potential for meeting the need for two-dimensional analysis. Benefits and limitations of alternative approaches, such as TEM staining and scanning probe microscopies, are also considered. Novel ion beam techniques are also extending ion beam analysis beyond depth profiling. Heavy ion backscattering is increasing the sensitivity for determining surface contamination, and nuclear microbeam techniques are being used to probe charge-collection efficiency in fine integrated circuit feature structures.

A. High depth resolution sputter profiling by SIMS

Secondary ion mass spectrometry is the only analytical technique available today that can be used to determine dopant-specific, in-depth profiles of atoms which are ion-implanted into materials at concentrations used for doping semiconductors. Electrical measurements such as spreading resistance can provide depth profiles, but only of carrier concentration, which does not distinguish between elements of a given charge type. In addition, these techniques measure electrically active species which makes it necessary to heat treat an ion implanted sample to activate the dopants prior to measurement. This treatment may move the implanted atoms from their initial depth locations which may defeat the purpose of an experiment to determine the in-depth profile of atoms as they are initially implanted into a material.

Secondary ion mass spectrometry has none of these drawbacks. Profiles can be obtained on unannealed or annealed specimens with equal ease and accuracy of the final result. This makes SIMS particularly attractive for calibrating process simulators of as-implanted distributions or in checking diffusion modeling using such programs as SUPREM III. Since the technique is dopant specific, modeling of diffusion of double implants of n-type dopants, such as As and P, can also be verified.

1. One-dimensional depth profiling

Implicit in the use of SIMS to calibrate process simulators is the ability to determine accurate in-depth distributions for dopants introduced into the very near-surface region of Si(2–100 nm). We will now discuss the recent advances in SIMS that have enabled the technique to measure dopant profiles in this regime. The need for near-surface profile determination is being driven by the ever-shrinking dimensions of Si devices which necessitates the use for ever-shallower junctions. This can be seen dramatically in Fig. 34 which compares an emitter-base junction from the early 1980s, [Fig. 34(a)], with a similar junction [Fig. 34(b)] from what we estimate devices will be like in the very near future. The emitter-base junction, which was at a depth of 1.65 μm 15 years ago, has moved up toward the surface so that now it is only a few hundred angstroms beneath it. In order to measure these shallow profiles correctly, the “surface transient” region of a SIMS analysis must be kept to a minimum. In addition, the very sharply falling emitter profile presents special difficulties for a SIMS analysis due to profile broadening by atomic mixing.

Atomic mixing results from the energetic interaction of the primary ion beam with the sample lattice atoms (Fig. 35). For SIMS, the most important type of atomic mixing is “cascade mixing” — which results from a primary ion striking a sample atom, displacing it from its lattice position so that it strikes adjacent atoms (which in turn may be displaced and strike other atoms) until the energy imparted in the collisions falls below the threshold for displacements. The net effect of this process is a general homogenization of all atoms within the depth affected by the collision cascade. Dopant atoms originally present at a given depth in the sample will be distributed throughout this “mixing depth” as sputtering proceeds so that the profile of the dopant will indicate a deeper distribution than is actually the case. It is thus of paramount importance that the depth of penetration of the primary ion into the sample be kept to an absolute minimum for high depth resolution profiling of shallow dopants in semiconductors. Referring to Fig. 35, it is obvious that the primary ion penetration depth is directly related to the mean projected range of the primary ion. This ion range can be reduced by the use of low ion energy and high ion mass.
However, for a given primary ion mean projected range, the primary ion penetration depth can be reduced if high angles of incidence are used (measured with respect to the sample surface normal).

The effects of primary ion energy and angle of incidence are shown in Fig. 36 for the analysis of a 30 keV As implant into single-crystal Si. This implant is expected to be used in the near future in fabrication of MOS field-effect transistors (MOSFETs). The broadening of the As profile is clearly visible when analyzed using high primary ion energies and lower angles of incidence. The SIMS analysis parameters indicated in Fig. 36 are not uncommon when profiling.

FIG. 34. (a) Diffused junction from the early 1980s. (b) Experimental diffused junction for technologies of the near future.

FIG. 35. Schematic of ion–solid interactions in SIMS.

FIG. 36. SIMS depth profiles of 30 keV As implant into Si analyzed with (1) a magnetic sector SIMS instrument using 14.5 keV Cs primary ion bombardment at 25° incidence, and (2) with a quadrupole SIMS instrument using 3 keV Cs primary ion bombardment at 60° incidence (from Ref. 132).
n-type dopants with current generation magnetic sector SIMS instruments and monitoring negatively charged secondary ions in order to obtain useable detection limits for P, As, and Sb.

A profile of an even more challenging sample is shown in Fig. 37(a). This example shows a SIMS profile of an implant of 5 keV arsenic in Si. The analysis was made using a cesium primary ion beam of 2 keV energy impacting the sample at an angle of 60° with respect to the surface normal. Negative secondary ions were detected. The Si profile in Fig. 37 shows that the traditional “pre-equilibrium” region of the profile is only 20 Å thick. This is a direct result of the shallow depth to which the Cs primary ions are implanted using these bombardment conditions. This narrow pre-equilibrium region allows us to accurately measure the As profile even though the peak concentration lies less than 70 Å beneath the sample surface. Figure 37(b) shows a boron depth profile taken from a Si sample implanted with BF₂ at an energy of 20 keV. This is equivalent to a B energy of only 4.5 keV. The profile was taken using the exact same instrument conditions to those used in the arsenic analysis shown in Fig. 37(a). Again, good definition of the B in-depth distribution is seen even though the peak concentration lies less than 200 Å beneath the sample surface.

The profiles in both parts of Fig. 37 show that both p- and n-type dopants can be analyzed in Si using the exact same high sensitivity, high depth resolution conditions. This leads naturally to the direct SIMS measurement of an actual ultrashallow p-n junction in Si. Such an analysis is shown in Fig. 38. The sample was an early attempt to form an
emitter-base junction of a poly-Si emitter NPN bipolar transistor by diffusing both B and As out of polycrystalline Si deposited on a single-crystal Si substrate. In a single measurement, both B and As were measured to define the junction, and oxygen was monitored to mark the position of the poly-Si/substrate interface. The analysis was made using 3 keV Cs primary ion bombardment at an impact angle of 60° with negative ion detection. Figure 38 expands the profile around the region of the junction. One can clearly measure the shallow diffusion of both dopants out of the poly-Si into the single-crystal Si substrate. The difference in diffusion coefficients between B and As causes a junction to form at a depth only 27 nm beneath the original surface of the Si substrate. The electric field effect of the interaction of the two dopants causes a “dip” in the boron profile which shows the location of the junction. The high depth resolution is confirmed by the fact that the As profile crosses the B profile exactly at the dip.

Figure 39 shows another example of our ability to measure ultrashallow junctions for materials that may be used in future technologies. The sample is single-crystal Si that was doped with boron using the gas immersion laser doping (GILD) technique. This doping technique forms an exceedingly sharp interface between the doped and undoped regions of the sample. If the boron profile shown in Fig. 39 was formed in a background of n-type doping at the three concentrations indicated, the junctions formed would be at depths of 15, 22, and 31 nm below the interface. The examples shown in Figs. 38 and 39 demonstrate the ability of SIMS as practiced today to measure ultrashallow junctions in Si which the SEMATECH and Semiconductor Industry Association (SIA) roadmaps for semiconductors predict will be needed as technologies shrink dimensions to less than 0.1 μm.

2. Two-dimensional depth profiling

The effective gate length of future devices may be less than 0.15 μm and it will be imperative to be able to measure the spreading of dopants under the gate with spatial resolution of approximately 0.01–0.02 μm. Direct cross-sectional SIMS measurements of dopants at the $5 \times 10^{16}$ cm$^{-3}$ level (1 part per million (ppm)) within these small areas will never be possible if one simply examines the number of atoms available. To obtain statistically useful signals, SIMS analysis requires 100 counts in the spectrum. To produce 100 counts, there must be $10^4$ dopant atoms in the analyzed volume assuming an ionization efficiency and instrument transmission combination of a factor of $10^{-2}$ (i.e., for 100 atoms, only one is detected). In a volume 0.02 μm on a side, there are only $2 \times 10^4$ Si atoms. For dopants at the $5 \times 10^{16}$ cm$^{-3}$ level there only 0.02 dopant atoms available for analysis. Several very specialized sample preparation techniques have been tried in an attempt to assess the degree of lateral spreading of dopants under mask edges. However, the emphasis in this area is clearly moving toward indirect methods of dopant distribution determination.

Indirect methods of measuring x–y dopant distributions (2D profiling) under mask edges and MOS gates are driven by the high spatial resolution requirements of this analysis. The general methodologies being tried are electron microscopy based and scanning probe microscopy based. A transmission electron microscopy based approach involves the
use of a selective chemical etch, the etch rate of which is dependent on dopant concentration. This produces a cross-sectional sample with a varying thickness which gives rise to contrast in the image. This contrast can be related subsequently to dopant concentration by calibrating the image contrast with dopant concentration in an unpatterned area of the sample using secondary ion mass spectrometry. An example of this method of 2D dopant determination is shown in Fig. 40. The example shows the TEM cross-sectional images and SIMS calibration depth profiles for two arsenic implants (35 and 120 keV) made into a CMOS structure in the gate region. The TEM image with selective etch clearly shows the spreading of the implanted species under the mask edge due to annealing. This imaging technique also shows that if the implant is angled towards the gate, the doping under it is increased (Fig. 41) and if the implant is angled away from the gate, the doping under is reduced (Fig. 42). Although impressive in their spatial resolution of dopant distribution, Figs. 40, 41, and 42 demonstrate the weakness of the selective-etch/TEM approach, i.e., its lack of contrast at doping concentrations of less than \(10^{19}\) atoms cm\(^{-3}\). Another electron microscopy based approach uses scanning electron microscopy\(^\text{139}\) in which the variation of secondary electron yield with dopant concentration is used to give a dopant-dependent image of the area of interest, again prepared in cross section.

Among the scanned probe microscopy based approaches being tried are (1) selective etch and atomic force microscope (AFM) measurement of surface relief\(^\text{140}\) and (2) the use of a scanning capacitance microscope (SCM) to provide dopant-dependent image information\(^\text{141}\). These techniques also require calibration of the measured signal using SIMS on an unpatterned region of the sample being measured in 2D. An example of the selective etch/AFM approach to 2D profiling is shown in Figs. 43 and 44 which show the topographic height produced in the sample by the dopant-dependent selective etch, the corresponding SIMS-measured dopant concentration (measured in an unpatterned area), and the reconstructed cross-sectional view of the dopant distribution near the gate edge of a MOS device. This technique shows somewhat greater sensitivity to low dopant concentrations than does the TEM approach, but it also poorer spatial resolution. An example of the SCM approach is shown in Fig. 45. The data were taken not on an actual MOS gate structure but, rather, on a test structure to show the ability of the technique to obtain dopant-dependent information on a submicrometer scale. Note the sensitivity to very low dopant concentrations, but also the apparent lack of sensitivity to differences in high dopant concentrations.

### B. High energy ion beam analysis techniques

The standard ion beam analysis method for characterization of ion-implanted or thin-film layers, such as silicides, is
backscattering spectrometry. This technique provides rapid (15 min), quantitative analysis of film thickness and depth profiles of major constituents. In the 1–3 MeV ion energy range, cross sections and energy loss are well known so that simulation programs can be used to characterize the samples. An example of a Rutherford backscattering spectrometry. This technique provides rapid (15 min), quantitative analysis of film thickness and depth profiles of major constituents. In the 1–3 MeV ion energy range, cross sections and energy loss are well known so that simulation programs can be used to characterize the samples. An example of a Rutherford backscattering spectrometry.

FIG. 43. AFM image of a calibrator sample after dopant selective etch, and vertical topographic line profiles through AA'. (a) Profile of calibrator after pre-etch step. (b) Profile of the calibrator after dopant selective etch. (c) Profile revealing the source/drain oxide of the calibrator after dopant selective etch. (d) Spreading resistance probe profile of the calibrator sample (from Ref. 140).

FIG. 44. Two-dimensional dopant profile of the calibrator sample. This map was created using a single AFM image and the calibration curve (a) from Fig. 43. Lines of constant concentration are delineated in intervals of 10^6 (from Ref. 140).

FIG. 45. Two-dimensional SCM micrograph of NIST test structure from a 1.3 μm×1.3 μm area around an ion implant edge, displayed with an alternating black/white color scale. The dark areas represent equi-C ref lines and are labeled with the respective chemical dopant concentrations from a correlated SIMS depth profile from an unpatterned region of the same sample (from Ref. 141).

FIG. 46. RBS spectrum from a magneto-optical rare-earth transition metal (RETM) multilayer thin film deposited on (001) Si.
analysis of a multilayer thin-film structure on Si is shown in Fig. 46. Film thicknesses or nonuniformities can be estimated quickly by the practitioner. The backscattering technique with 1–3 MeV He ions is in use in several hundred laboratories worldwide due to the availability of MV accelerators. The technique is well established as an analytical tool in materials science studies with descriptions in undergraduate/graduate textbooks.\textsuperscript{144,145}

Recent developments that improve the capabilities of ion beam techniques will now be described. The developments range from use of high energies (\(E > 3\) MeV), medium energies (100–300 keV), time-of-flight (TOF) detection systems, and microbeams, all designed for specific applications.

1. High energy (\(E > 3\) MeV) analysis with He ions

At energies above 3 MeV, resonances\textsuperscript{146} appear in the He scattering cross section as shown for oxygen and carbon in Fig. 47. These resonances lead to enhanced backscattering yields. These resonances are now used routinely at energies between 3 and 5 MeV to obtain depth profiles of O and C in a variety of samples such as high \(T_c\) superconductors\textsuperscript{147} and SiGeC layers on Si.\textsuperscript{148} This extension of backscattering to higher energies is now accommodated in simulation programs.

High energy tandem accelerators with terminal voltages greater than 4.0 MeV, are available and provide a source of high energy heavy ions. These ions, such as 20 MeV Si\textsuperscript{5+}, are used in the elastic recoil detection (ERD) mode of forward scattering for depth profiling of light mass ions in medium mass substrates such as Si.\textsuperscript{139} The scattering cross sections are Rutherford and energy loss values are known\textsuperscript{146} so that reliable concentration profiles can be found. An example of the analysis of carbon in a chemical vapor deposited layer of SiGeC on Si is shown in Fig. 48.

By using even heavier ion beams, such as 130 MeV \textsuperscript{127}I or 200 MeV \textsuperscript{197}Au, ERD becomes essentially a universal analytical technique, capable of analyzing simultaneously all elements from H up to mass 100 or even higher, with almost constant detection sensitivity.\textsuperscript{151,152}

2. Medium energy (50–500 keV) for high sensitivity analysis

The ability to measure very small levels of impurities on surfaces is increasingly important in many areas of technology, particularly in microelectronics, where integrated-circuit device sizes become ever smaller and as few as \(10^{10}\) cm\(^{-2}\) heavy-metal impurities can create problems. To achieve this sensitivity, lower energies and heavy ion backscattering spectrometry (HIBS) are used.\textsuperscript{153} The backscattering yield is proportional to the square of the atomic number of the analyzing beam and inversely proportional to the square of its energy, so an enhanced yield can be obtained by using a higher Z ion beam at lower energy. Although this yield enhancement is well known, it is not widely exploited because in most situations pileup in surface barrier detectors (SBDs) due to yield from the lower mass substrate overwhelms the signal from ions scattered by trace surface impurities.

The TOF spectrometer is used for high sensitivity detection of impurities on solid surfaces with medium energy ions.\textsuperscript{154,155} The spectrometer (shown in Fig. 49) measures the interval between the detection of secondary electrons emitted...
as a particle emerges from a thin carbon foil placed in its path and the subsequent impact of the particle on a microchannel plate. The spectrometer does not have to be scanned and is sensitive to all charge states of the particle; its throughput can be considerably larger than that of devices such as electrostatic analyzers. Time-of-flight detection systems for medium energy, heavy ion scattering are being developed for hydrogen detection (forward recoils) and for surface contaminant control (backscattering HIBS). The most sensitive of existing general-purpose tools, total reflection x-ray fluorescence (TXRF), can detect $\sim 1 \times 10^{10}$ atoms cm$^{-2}$ levels of some elements such as Fe and Cu, but for many others it is limited to $1 \times 10^{12}$ atoms cm$^{-2}$ or worse. Although a sensitivity of $10^8$ atoms cm$^{-2}$ is routinely claimed for TXRF, this was achieved only via preconcentration through a technique called vapor phase decomposition, wherein an acidic mist is applied to the wafer surface and then concentrated to a small spot. Obviously, this technique has many uncertainties. HIBS provides a new ion beam analysis capability with the potential for providing the required sensitivity without pre-concentration for all elements above Ar. With preconcentration the sensitivity would be approximately two orders of magnitude higher. Estimates of the detection limits for contamination are shown in Fig. 50. Sputtering of the sample surface by the analysis beam is the ultimate limit to the statistics that can be obtained, and hence the sensitivity. The rate of sputtering depends on the state of the impurity: particles sputter differently from a thin uniform layer than from a widely dispersed, low concentration layer so it is difficult to predict the limit in advance. These experiments suggest that the rate for a submonolayer impurity is lower than predicted by a simple sputtering model using bulk sputtering rates.

### 3. Microbeams

The micrometer-size ion beam has arrived, with its own conferences, the International Conference on Nuclear Microprobe Technology and Applications (Shanghai, 1994; Albuquerque, NM, 1996), and with 50 nuclear microprobes worldwide and a growth of about 2 per year. Most labs use their microprobes for analytical applications and all of the ion beam analysis techniques have now been applied to nuclear microprobe analysis. Some ion beam analysis techniques have actually been initially used on the microprobes, including transmission ion microscopy, elastic scattering analysis, and ion beam induced luminescence.

The figure of merit for microbeams is to get 100 pA into a 1-μm-diameter spot. By simply apertureing the beam before it enters the lens, virtually all of the lens aberrations can be eliminated, at a severe cost in beam current, and beams of 0.1 μm have been obtained in several labs. Most lens and quadrupole systems have been installed on 2–3 MV accelerators. The size of the microbeam and, consequently, the beam current determine which types of analytical techniques can be performed. To perform routine analytical techniques such as RBS requires 1–100 pA, which limits the minimum beam size to 1 μm. An example of the use of a 1.5 MeV helium ion microprobe (4 μm×4 μm) with Rutherford backscattering to study the lateral growth of cobalt silicide on Si is shown in Fig. 51. In self-aligned silicide (salicide) structures, the extent of lateral silicide growth (several μm) is a critical issue. Using the microprobe, lateral and cross-sectional cobalt distributions in cobalt silicide patterns could be directly observed. Comparison between arsenic-implanted and unimplanted patterns revealed that ion beam mixing by arsenic implantation suppressed the lateral growth of cobalt silicide during rapid thermal annealing at 1000 °C for 1 s.

Larger microbeams [on the order of 20–40 μm of doubly charged helium at 1.5 MeV (Ref. 157)] have higher beam currents of 3–10 nA. Although at this resolution it is impossible to directly image device features, this type of system is being used for analysis of aluminum alloy films in wiring metallization and wire bond pads, and in measuring silicide stoichiometry in test structures and large area features used in packaging applications.

In addition to these routine analytical techniques, new techniques that require only fA of current have been invented...
and they are ideal for application with the 0.1 \( \mu \text{m} \) beams (which have small currents). These are the single ion effect techniques. The applications involve examination and modification of integrated circuits (ICs), single biological cells, exposing photoresists, and making tracks (i.e., Coulomb explosions) in materials. For ICs two techniques were invented: single event upset (SEU) imaging and ion beam induced charge collection (IBICC).159,160

SEU imaging is a diagnostic technique which directly images, with \( \mu \text{m} \) resolution, those regions within an integrated circuit that are susceptible to ion-induced malfunctions. Such malfunctions are an increasing threat to space-based systems which make use of current-generation IC designs. A complementary technique to SEU imaging involves measurement of the charge collection volumes within integrated circuits; IBICC has been used to generate \( \mu \text{m} \) resolution maps of the charge collection response of integrated circuits. Combining the SEU imaging and IBICC techniques provides a better understanding of single event upset phenomena. High resolution IBICC images are used to extract more information from charge collection spectra than that obtained from conventional broad-area ion exposures, such as from radioactive sources. As reductions in circuit feature size continue in the submicrometer regime, IBICC could prove to be a technologically valuable replacement for electron beam induced current (EBIC).159,160

C. Characterization status and future directions

Great advances in high depth resolution in dopant profiling using secondary ion mass spectrometry have taken place in the last fifteen years. Junction depths of less than 20 nm can be measured today. The current state of the art in one-dimensional dopant distribution measurement is sufficiently advanced to carry us well into the next century of semiconductor processing advances.

What cannot be measured directly today using SIMS is the spatial, two-dimensional distribution of dopants with 10 nm spatial resolution. This type of data is needed to calibrate process simulators in order to model the lateral spreading of dopants under mask edges in the critical gate region of MOS transistors as gate lengths shrink. It appears that indirect imaging methods will be used to obtain this information, with various methods now being tried for obtaining dopant concentration-dependent contrast in those images.

Ion beam analysis techniques have been used in conjunction with ion implantation for the past thirty years. Some techniques such as Rutherford backscattering spectrometry with 2 MeV \(^4\)He ions and mm beam area are well established as routine analytical tools. Improvements are going beyond standard MeV He ion backscattering. One such area is the reduction in beam size to submicrometer dimensions (micro-beams). Other areas are (1) use of higher energies (3–10 MeV) for resonance analysis of C, N, and O; (2) use of heavier mass ions (at high energies) for elastic recoil detection of low mass ions; and (3) use of medium energy ions and time-of-flight detection for high sensitivity surface analysis. The development of time-of-flight detection systems grows apace. Ion beam analysis continues to be a field where new approaches are developed for special analysis problems.
active investigation the challenges for scaling of shallow junctions into Si with a beamline technology. In the following, we break up or solid-source diffusion with rapid thermal annealing ~RTA! to fabricate fully functional CMOS transistors with 50 nm gate dimensions along with heavily doped, Ti salicide-capped lightly doped drain ~LDD! and channel-pocket implants along with heavily doped, Ti salicide-capped source/drain implants. Although many other techniques are under active investigation ~plasma immersion implantation, vapor or solid-source diffusion with rapid thermal annealing ~RTA!, gas-immersion laser doping!, the dominant method of doping remains the direct ion implantation of dopants into Si with a beamline technology. In the following, we break up the challenges for scaling of shallow junctions into ~1! development of technology for production-worthy, low-energy ion beam tools and ~2! control of materials issues such as dopant diffusion and activation and defect annealing with low thermal budgets.

V. SHALLOW JUNCTIONS: MACHINE AND PROCESSING CONSTRAINTS

Processing capabilities are ultimately limited by the ability to develop machines that achieve these goals on manufacturable scales. It has been suggested that the continuing growth of the electronic industry will be determined by manufacturing efficiency rather than by the traditional technological factors of decreased feature size, increased die size, and yield improvements. Since scaling of Si IC devices towards 0.1 mm gate dimensions requires shrinkage of source/drain junctions depths to ν0.1 mm and control on lateral dimensions to ±10 Å,12 advances must be made to economically produce shallower implants with greater control of lateral features. The general trends for CMOS transistor feature sizes and junction locations have been modeled extensively by SEMATECH161 and are incorporated in the recent SIA roadmap for semiconductor technology.162 By comparing the source/drain junctions and the depth of peak doping for CMOS wells to the mean range of ions in Si ~Fig. 52! one can get an overview of the energy requirements for ion implantation tools for each new generation of devices.

There are no widely recognized fundamental barriers to fabrication of CMOS devices as small as 10 nm in gate length. In 1994, a standard production implanter was used to fabricate fully functional CMOS transistors with 50 nm gate dimensions ~and 3.1 nm gate oxide! using scaled spacer lightly doped drain ~LDD! and channel-pocket implants along with heavily doped, Ti salicide-capped source/drain implants. Although many other techniques are under active investigation ~plasma immersion implantation, vapor or solid-source diffusion with rapid thermal annealing ~RTA!, gas-immersion laser doping!, the dominant method of doping remains the direct ion implantation of dopants into Si with a beamline technology. In the following, we break up the challenges for scaling of shallow junctions into ~1! development of technology for production-worthy, low-energy ion beam tools and ~2! control of materials issues such as dopant diffusion and activation and defect annealing with low thermal budgets.

A. Technology of ion implantation beamlines

1. Space-charge balance effects

Many of the technology challenges for low energy ion implantation for doping of shallow junctions are associated with the character of the balance of charged species in the ion beam plasma. For the almost universal case of positive ion beams, the space-charge levels are balanced by electrons and other positive ions in the region of the ion beam. The degree to which the effect of the beam space charge is controlled by machine design and operation determines ~1! the beam transmission and beam current on the wafer ~and through that the tool throughput and economic productivity!, ~2! wafer charging levels ~affecting both dosimetry and damage to thin dielectric layers!, and ~3! beam divergence ~beam incidence angles on the device surfaces and topography!.

2. Transport of high beam currents

The main focus of designs for low energy implantation has been to increase the beam current on the wafer for B ion beams. The challenges are ~1! to develop stable, low-noise, high-density plasmas in the ion source, ~2! to extract ions from the source with high efficiency and adequate beam focusing, and ~3! to transport low energy beams over a distance of the order of 2 m ~through analyzing magnets, mass-resolving slits, focusing, and charge-control assemblies! to the wafer.

The principal difficulties are in the design and operation of the extraction electrodes and in maintaining sufficient control on the space-charge balance in the beam to be able to transport a high current, low energy beam of ions from the ion source to the wafer. As shown in Fig. 53, the increase in the beam divergence of a 95% space-charge neutral boron beam as the energy is reduced below 10 keV ~Ref. 163! is comparable to that of an uncompensated positive ion beam as the beam current is increased from μA to mA. Space-
charge limits to available extraction current are also more severe as the energy of the ions is reduced. This limit is degraded further by instabilities in the ion source plasma, fluctuations in the extraction power supply, and strong defocusing effects associated with the curvature of the boundary between the ion source plasma and the extracted ion beam region.\textsuperscript{165} The path to progress, in the face of these many interacting effects, has been highly empirical and only in recent years has a number of useful plasma modeling programs been available for use by the implant tool developers.\textsuperscript{164,165}

Significantly higher beam currents can be transported by extracting ions at a high voltage, transporting these relatively high energy ions through the beamline, and then decelerating the ion beam to the desired energy in a region just before impact with the wafer. For various combinations of voltage ratios and beamline designs, the increase in the beam current on the wafer for accel/decel operation can be 5–10 times higher than in “drift” mode at low energies. The penalty for this approach is that any ions that are neutralized in the beam after the mass analysis stage are not affected by the decelerating electrodes and arrive at the wafer at the energy of the extraction stage. Under these conditions, the high energy, neutral fraction of the incoming ions penetrates deeper into the Si than the intended profile and also is not usually measured by the tool dosimetry, causing a dose overrun.\textsuperscript{166} The key to developing high productivity accel/decel beams without any detrimental effect by neutral beam components is a better understanding of the vacuum effects in the beamline and knowledge of the energy and ion dependence of the neutralization cross sections for dopant ions.

### 3. Dose uniformity

Instabilities in the beam profile are a problem in beamline designs where there is a relatively long open path to the wafer surface from the last active focusing element and there is an insufficient supply of low energy electrons to balance the positive ion charge. Small positive charge buildup on the wafer surface, leading to surface potentials of a few tens of volts, can rapidly deplete the electrons in the ion beam plasma leading to “blowup” of the ion beam as it crosses the wafer surface.\textsuperscript{167} This gives a “bulls-eye” pattern to the dose distribution (Fig. 54). Solutions to this problem have been to provide additional sources of electrons to the ion beam through various “electron shower” devices or to design beamlines with strong focusing elements that can control the beam size near the wafer surface. However problems with beam size instabilities driven by space-charge balance problems continue to be reported, especially at lower beam energies.\textsuperscript{168}

### 4. Surface charging

The principal driving force behind the increased concern for wafer charging effects is the shrinkage of the thickness of dielectric layers, especially gate “oxides,” as feature sizes are scaled down (see oxide thickness trends in Fig. 52). For increasingly thin dielectric layers, the surface potentials, generated by surface charge imbalances on the wafer, that can be tolerated decrease roughly proportional to the oxide thickness.

The principal charged species that play a role in the interaction of the beam plasma with the device structures on the wafer surface are (1) the energetic dopant ions, (2) ionized atoms (“slow ions”) which are generated by collisions between the energetic dopant ions and background gas atoms, energetic ions, (3) electrons generated by collisions of the dopant beam ions with background gas atoms and secondaries generated from ion bombardment of the wafer and machine surfaces, and (4) charged particles (ions and electrons) supplied from various charge control devices (electron showers or plasma sources) in the implantation tool.\textsuperscript{169,170} Additional charge flows that can play a role are (1) charged atoms that evolve from the outgassing of materials from photoresist layers, (2) secondary ions generated on surface structures, and (3) unusual conduction paths on the wafer surface, such as surface layers of ion bombarded oxides.

The real challenge for control of wafer charging is to not only avoid catastrophic breakdown events, such as those that occur when the surface potential exceeds the breakdown field, but to avoid degradation of gate dielectrics caused by in-process current flows through the dielectric that can show up later as poor chip reliability. Some of the leading concerns for reliability are (1) population of process-induced oxide damage sites by “hot carriers” in the gate oxide near the CMOS–drain junction (leading to threshold voltage shifts as the device operates), (2) dielectric leakage in DRAM devices (leading to charge loss in the storage cells or the need for more frequent refresh cycles), (3) degradation of electronically erasable programmable read-only memory EEPROM devices (limiting the number of useful memory store operations).

A major recent advance in the understanding of charging effects is the ability to measure the current flows and voltage conditions on wafer surfaces exposed to ion beam processing.
plasmas with EEPROM-based sense and memory structures. EEPROM-based sensors give direct measures of the \( J-V \) characteristics of the ion beam plasma, without the need to deconvolute the data through the understanding of oxide wearout and breakdown phenomena. The example shown in Fig. 55 is a map of the surface potential variation over a wafer exposed to a "native" 16 mA As ion beam with no charge control techniques in operation.\(^{169}\) The high surface potentials (\( \approx 17 \text{ V} \)) seen near the edges of the wafer where the ion beam enters and leaves the wafer surface are reduced in the middle portion of the wafer to \( \approx 10 \text{ V} \) by the effects of secondary electrons from the Al sensor pads. Analysis of the flux–voltage characteristics of the ion beam plasma shows that the lowering of surface potentials in the center portions of the wafer was accompanied by an increase in the plasma density over the wafer surface to \( \approx 100 \times \) the ion beam density. Consideration of these results has led to the increased interest in "cold plasma" sources for control of wafer charging in advanced ion implantation systems.\(^{171}\)

A comprehensive understanding of the mechanisms that lead to device damage during exposure to ion beam processing plasma is critical for ion implantation, plasma etching, resist stripping, and sputtering technologies. As dielectric layers are thinned as device features shrink, both the tolerable surface electric fields and surface currents are driven to the limits of detection of even the new EEPROM-based sensors. Many of the measurement and modeling tools that are needed to achieve this comprehensive understanding are not available, or are in need of considerable enhancement. There is a particular need for the modeling of the interaction of actual device structures with ion beam processing plasmas.

5. Beam divergence effects on lateral junction location

For 0.25 \( \mu \text{m} \) and smaller devices, the requirements for control of the levels and location of doping profiles exceed even the average separation between dopants for the lightly doped drain junction (for \( 3 \times 10^{17} \text{ atoms cm}^{-3} \) doping levels, the average dopant separation is 15 nm, whereas the required control on the location of the LDD junction is estimated to be 5 nm).\(^{172}\) Given the strong sensitivity of critical transistor parameters such as threshold voltage, subthreshold leakage current, and circuit speed (Fig. 56) with junction doping profiles, a few errant atoms can produce significant effects! In these situations, the precision to which wafer beam orientation and beam divergence is controlled will be critical to the control of lateral doping profiles under mask edges. The risk for beam divergence variations increases as the ion beam energy decreases due to the space-charge balance problems of the sort discussed earlier for beam spot size controls. While beam spot size fluctuations which give rise to wafer-scale dose variations can be controlled by proper beamline design and beam space-charge controls,\(^{167}\) the impact of beam divergence variations, perhaps even on the scale of the device structures, has not yet been sufficiently characterized.

B. Materials issues linked to machine design and process

1. Pre-amorphization process and enhanced diffusion

Two key materials issues that affect shallow junction formation are (1) the role of channeling on as-implanted dopant and damage distributions and (2) control of damage-enhanced dopant diffusion and understanding of its relationship to damage accumulation during implantation and defect evolution during annealing. The materials aspects of these issues (TED, etc.) are discussed in earlier sections of this review. On the IC fabrication process side of these issues, the debate continues on the practicality of preamorphization implants to suppress channeling with Si, Ge, Sb, or In beams and whether the complications introduced by these processes are worth the effort in the face of the strong effects of transient enhanced dopant diffusion. An example of this issue is shown in Fig. 57 for the case of B implants at energies from

FIG. 55. Map of the surface potentials on a wafer populated with charge monitoring sensor arrays after exposure to “native” 16 mA As, with no charge control techniques in operation (from Ref. 169).

FIG. 56. Variation in subthreshold leakage currents with source/drain junction depth for 0.25 \( \mu \text{m} \) transistors (note changes of \( 200 \times \) in leakage for 10% changes in junction depth) (from Ref. 172).

FIG. 56. Variation in subthreshold leakage currents with source/drain junction depth for 0.25 \( \mu \text{m} \) transistors (note changes of \( 200 \times \) in leakage for 10% changes in junction depth) (from Ref. 172).
2.5 to 10 keV into Si(100) materials and Si that was amorphized prior to the B implant by Ge ion implantation. While the Ge implant resulted in shallower B implanted profiles, the final junction depth after a short anneal cycle was close to the junctions implanted into crystalline Si.

Nevertheless, the wide use of preamorphization implants for high dose, low energy boron processing for CMOS source/drain junctions will put new requirements on this “bottleneck” process step. Machine technology improvements such as rapid beam setup automation (to reduce the 5–10 min delay per beam change in present technology to less than 1 min) will be of great value in this area. In addition, the use of other novel implantation procedures, such as implantation at temperatures far from the room ambient or the use multispecies and multistep implants, can have a profound effect on new implantation system design.

2. Damage accumulation for low energy boron in Si

The damage accumulation for low energy B implants is particularly important since, for beam energies of less than ~10 keV, the damage and stopping mechanisms are now dominated by the nuclear, or elastic, stopping power rather than by the electronic stopping. This is a new regime for damage effects of B in Si implants. Effects of beam current and wafer temperature on dopant profiles, damage accumulation, and dopant activation, which have been studied under “research” conditions, are now encountered under “industrial” conditions for 0.5 and 0.35 μm device fabrication.173

For low energy B implants, the general trend is that, as the beam current is increased, the damage accumulated in Si at a given dose increases. After annealing, the more highly damaged material often shows a higher degree of dopant activation.173 The increased damage accumulation is seen in higher Thermo-wave probe photoacoustic signals for the as-implanted Si; the dopant activation increases are seen in a lower sheet resistance after short time anneals for high beam current implants. The complexity of this situation is far from fully understood. For instance, the damage accumulation has been seen to depend on the levels of pre-existing damage in the Si materials before the low energy B implant. The existence of damage from a preimplant with Ge at a dose of 10^{14} Ge cm^{-2} greatly increased the damage levels for 2.5 keV B implants compared to a lighter Ge predamage of 5 \times 10^{13} Ge cm^{-2} or implants into undamaged Si(100).

3. Defect engineering

The goal of defect engineering work is to find conditions where there is no residual damage remaining in implanted Si after modest thermal treatments. In addition to studies of the effects of beam current and wafer temperature other approaches that have been explored to some extent are (1) the use of multiple-species implants for control of dopant diffusion and defect annealing,174,175 (2) the use of cold wafer temperatures (77–250 K) during implantation to increase damage accumulation to above amorphization levels,176,177 and (3) the use of implant-annealing sequences made up of partial implant doses and short thermal anneals.69,178

Studies of the use of nondopant ions in Si has centered on predamage implants to suppress channeling of the dopant ions (particularly B) and on creation of gettering sites in the implanted layer to reduce metallic contamination levels in the device region. However, nondopant species such as F can have a strong effect on B diffusion (F retards B diffusion (Fig. 58),174 and residual defect levels after annealing.175 Implanted F is trapped at defect layers, in F-filled voids and at high dopant concentrations (Fig. 59).175 The amount of trapped F is strongly dependent on the initial F concentration and the anneal temperature. Since F forms a strong F–Si bond, it is thought that F competes with B for interactions with Si interstitials created by the implant damage process.
Since B–Si interstitial pairs are a key species in the rapid diffusion of B at the early stages of annealing, one effect of the presence of mobile F is to reduce the available population of free Si interstitials and thereby reduce TED effects for B. Even partial success in this area will have a strong impact on the design and operation of implantation machines by increasing the need for flexible ion source and beamline technologies that can deliver useful quantities of nondopant species with production-level reliability and purity in addition to the cost-effective processes for dopant species.

On the extreme end of the need for control of wafer temperature to stabilize damage accumulation levels at near-room temperatures is a study of the use of refrigerants and special "cryo" techniques to achieve wafer temperatures during implantation at below 0 °C under some approximation of production conditions. The need to maintain stable wafer temperatures at below-room temperature calls for large refrigeration systems, heat-flow baffles, cold-resistant seals, and elastomers, and high-vacuum conditions. The enhanced vacuum requirement is to prevent the wafers from absorbing background gases and masking the ion beam. Although several production-class tools have been built to operate in the −30–0 °C range, the net benefit to device processing is not yet clear.

In a recent study, the damage threshold for formation of stable defects (dislocations) after annealing was mapped out for implants into Si at various temperatures and at complantation conditions. The damage threshold for B in Si is an order of magnitude lower than that for As and Sb ions. The energy range surveyed in these studies ranged from 50 keV to 2 MeV. It would clearly be valuable to extend these studies into the range of shallow-junction formation (<10 keV for B). A consequence of these findings is that damage-free material could be formed by a sequence of partial dose implants and intermediate anneal steps. This would introduce a radical new view of "batch" processing. For this process to be implemented on an industrial scale, the design of tools that could perform rapid partial implants and then anneals in sequence for each wafer would be very different from what is presently in use or what is routinely projected for future requirements.

4. Dose control issues

For shallow junction implants, especially for heavy ions such as As or Sb, the density of the target surface increases markedly during the course of the implant. This density increase changes the net stopping power of the dopant-rich Si so that the dopant ions arriving during the later stages of the implant have a shallower profile than do the early ions. Another significant effect is the loss of dopant atoms from the implant profile due to sputtering of the surface. This phenomenon, a routine effect in implantation into GaAs, has been reported for 6 keV As source/drain implants. For 6 keV As in Si, the sputter-limited saturation dose is $2 \times 10^{16}$ As cm$^{-2}$ which is a factor of 10 above the dose required for source/drain doping for 0.25 μm transistors. However, this effect, combined with increased ion reflection, presents a fundamental limit to the net dose that can be implanted into shallow junctions in Si (particularly for heavy ions) which may be an issue for novel device designs.

Dose uniformity and lot-to-lot dose repeatability have long been among the chief advantages of doping by ion implantation. The development of mapping techniques for wafer-scale images of dose monitor data has led to rapid improvements in ion implanter design and these techniques are routinely used to identify and diagnose dose control malfunctions. However, the increases in wafer sizes have increased the area over which the ion dose must be distributed. This challenge is particularly strong for single-wafer systems such as those routinely used for lower dose implants. For batch-style implanters, the relevant area is the batch scanned area, which is less sensitive to wafer size since it is proportional to the wafer diameter (in most designs) rather than the wafer area.

A much stronger driver is the dramatic increases in dose and profile control that are required to fabricate CMOS transistors for gate lengths of less than 0.5 μm. Modeling of 0.25 μm transistors show that a 10% shift in the source/drain junction depth (a shift 300 Å in this model) results in an increase in the subthreshold leakage current by a factor of 200. Extensions of the historical trends in dose uniformity requirements also predict real needs for dose control approaching 0.1% for critical applications of 0.25 μm technology. The start of 0.25 μm production is forecast for 1998, the same year as the introduction of 300 mm wafer fabs. The combination of increased wafer area and tightened dose controls results in a four order of magnitude tightening of the "quality factor," the dose uniformity error (1 sigma) divided by the wafer area (Fig. 60).

Dose control at the level of 0.1% is a significant challenge for present-day technology for both implantation systems and metrology. Although tightly monitored implantation systems can now deliver routine dose uniformity in the range of 0.25%, these same systems are usually specified at 0.5% or even 0.75% levels. Injudicious process choices and system malfunctions can result in much higher dose variations.
errors. A closely related requirement is the need for improved metrology for measuring dose to accuracies of better than 0.1% over the complete process dose range (5 \times 10^{10} - 5 \times 10^{16} ions cm^{-2}). Present methods, based on four-point-probe sheet resistance measurements or modulated laser pulse-induced reflectance effects, cannot achieve this at any dose or energy condition and are often an order of magnitude less sensitive than this 0.1% goal. These dose measurement metrologies are also sensitive to damage accumulation and room temperature annealing effects, a serious problem for process control and system calibration. An ideal dose metrology should also be able to monitor the delivered dose either during the actual implant or at least in situ, directly after the completion of a single implant cycle.

An additional challenge for dose control is the increasing need to maintain dose uniformity on the scale of an IC die and down to the separation distance between active devices. This small-scale doping control becomes increasingly critical as the die size increases. For small die sizes (0.5 cm or less), dose fluctuations on the scale of several cm would result in variations in chip performance, resulting in a problem of die sorting and various levels of conformance to the IC specifications. But for a die size on the scale of a 64M DRAM or advanced \( \mu \)-processor chip, \( \approx 2 \) cm, that same dose fluctuation can result in zero yield for all die due to the cumulative effects of variations in transistor speed, gain, and noise margins across the chip causing loss of the synchronization of signals at critical nodes.

For any beam-scanning technology, the key to good micronuniformity is to achieve a high degree of overlap between adjacent passes of the ion beam. The required degree of overlap is higher than is commonly thought; for a “simple,” square profile beam to achieve a better than 1% level in local dose fluctuations, the distance between the centerline of the beam on adjacent passes must be closer than 10% of the beam width [full width at half-maximum (FWHM)]. For more complex beam shapes, the effects of local “hot spots” and sharp beam current density gradients all contribute to degradation of the dose uniformity on the microscale. Increasing the number of passes the beam makes over the die, although it helps to reduce the effects of single-event beam current fluctuations (hashes, ticks, and dropouts), has only a marginal effect on the micronuniformity levels determined by the beam size and scan overlap.

In the world of 0.25 \( \mu \)m devices and smaller, where “every atom counts,” any mechanism that contributes a nonstatistical dose fluctuation must be eliminated. Here again the challenge is for implant beamline and scanning system design, dose metrology on the scale of micrometers to centimeters, and astute choices for process conditions.

### C. Alternative doping methods for shallow junctions

The critical nature of the search for cost-effective technologies for advanced IC processing has stimulated the development of many innovative approaches to find an alternative to the beamline implanters, the technique of choice for doping of CMOS devices at the 0.25 \( \mu \)m scale and larger. These innovations are coming from a wide mix of industrial research by equipment manufacturers, university research, and (in Japan) government stimulated programs. A brief discussion of these alternative doping techniques follows.

#### 1. Plasma immersion ion implantation

Plasma immersion ion implantation (PIII) is the most mature of the alternative technologies and has been actively investigated for semiconductor fabrication and metallurgical application for nearly a decade. Since this technique is a direct extension of ion implantation processing, work on the materials issues for PIII can build directly on the wealth of understanding derived from implantation with beamline machines. The principal advantages of PIII are compact system size (by elimination of acceleration and mass resolution components used in beamline systems) and high efficiency for low energy, high dose implantation, a critical “bottleneck” for standard implantation processing. Progress towards industrial use of PIII has been reported in a number of workshops.

#### 2. Vapor-phase and CVD doping (thermal doping)

Doping by deposition and diffusion from vapor-phase dopants was the technique that ion implantation originally replaced in the semiconductor fabrication line. It has been reinvestigated for use for shallow junctions. The principal advantages of thermal doping methods are that they do not need to consider such ion implantation issues as channeling, lattice damage, and wafer charging. However, these methods are strongly limited by the substantial thermal budget required to introduce dopants into Si by surface reactions and solid-state diffusion. Thermal doping methods also require use of “hard” masks, such as chemical vapor deposition (CVD) oxides, rather than the polymer-based photoresists used for masking in ion implantation. The process complexity increase needs to deposit, pattern, and remove these hard mask layers for each thermal doping step is a substantial barrier to use of these methods in routine Si device fabrication.
3. Gas-immersion laser doping

The thermal budget and masking issues of thermal doping are eliminated by the use of pulsed-laser power to melt selected Si regions in the presence of gas-phase dopants. Since the diffusion of dopants in GILD is through liquid-phase Si, the dopant motion is extremely rapid. Since only selected surface regions are melted, the rest of the Si wafer is not substantially heated, which eliminates issues such as wafer warpage by conventional diffusion furnaces and rapid annealing systems. The GILD approach could lead to substantial process simplification when it is coupled to a projection masking system so that an entire sequence of photore sist application and patterning, ion implantation, resist removal, cleaning, and annealing are all replaced by a single GILD step.

4. Atomic-layer epitaxy

The application of molecular-beam epitaxy (MBE), techniques to doping of Si has produced some remarkable precisely controlled doping distributions. In combination with Si epitaxy, doping layers of a single atomic diameter (“delta” doping) has been achieved. However these techniques face similar thermal budget and masking limitations as more conventional thermal doping processes and also require extraordinary levels of surface cleanliness and vacuum purity.

5. Cluster beam implantation

A new approach to ultralow energy (100 eV–1 keV) implantation is to form beams of ions that contain a large number (10−102) of atoms. If these clusters of atoms are ionized to a low charge state and accelerated, the kinetic energy is shared between the atoms in the cluster. For instance, if a cluster containing 10 atoms is ionized to a +1 charge state and accelerated across a 1 keV potential, the average energy per atom is only 100 eV. The large mass of the accelerated species strongly reduces the space-charge balance problems for single atoms in beams at low energies.

While many of the techniques for beam transport and wafer handling can be directly applied to cluster-beam implantation systems, the materials aspect of the process is significantly different from implantation with single atom beams. When ions arrive in a beam of single atoms, the individual stopping events are complete in a time (~10−12 s) that is much shorter than the average arrival rate of single ions (an ion every ~10−6 s for beam current densities of ~1 mA cm−2). When a cluster of atoms arrives at a surface, the stopping events occur for all of the atoms in the time scale of 1–20 ps. On the atomic level, the process results in a local melting of the Si crystal with shock wave-driven displacements of atoms in the lateral and vertical directions. One aspect of the strong collective motions that occur on impact of a cluster with a solid surface is that, unlike the roughening that occurs for high dose implants with single atoms, impacts with beams of cluster ions result in a reduction in the surface microroughness. The damage accumulation rate for cluster-beam implantation is expected to show dramatic reductions when the average energy per atom is reduced to the threshold energy for lattice displace-ments (~15 eV for Si), opening another path towards the goal of “damage free” implantation processing.

D. Additional considerations: Large-area substrates, SOI, modeling requirements

In addition to the challenges outlined in Sec. V, there are other forces that could significantly change the process and machine aspects for doping of shallow junctions. On the machine technology side, the continued trend towards larger Si wafers (with 300 mm fab lines expected to start up in 1998) will drive the design of wafer handling mechanisms towards the conditions needed to process large-area flat-panel devices. Trade-offs in the complexity of these wafer handling systems and beamline issues such as scanning rates and beam size control will determine the design criteria for the next generation of implantation tools.

On the materials side, the widespread use of SOI wafers would bring a dramatic change in doping process requirements. In SOI materials, the deep-implanted well doping steps are eliminated, bringing about a reduction in the number of implantation steps for CMOS from 15–20 to the order of 5–8. Since the junction depth for shallow junctions is determined by the thickness of the surface Si layer (~500 Å for thin-layer devices), the need to control ion channeling and diffusion rates is eliminated or significantly reduced. SOI wafers also present significant process advantages for such advanced techniques as PIII and GILD since both approaches provide lower (PIII) or no (GILD) sensitivity to wafer charging effects.

The extreme levels of control of doping profiles throughout the volume of small-scale devices in order to obtain functional IC devices will continue to transform the process development cycle. Advanced process development will rely on increasingly sophisticated simulation and characterization tools with 3D capabilities. Requirements for elimination of surface contaminants from Si processing is already approaching the ppm of a monolayer level (1010 atoms cm−2) and will certainly continue towards lower levels.

VI. LITHOGRAPHY AND FOCUSED BEAMS

In addition to the increased need for control of implant depth, greater lateral control of dopant profiles is also required. In the following Sec. VI, we focus on three methods of using ions for creating patterned doses, shown in Figs. 61(a)–61(c) focused point beams, proximity masks, and stencil mask image projection, respectively. Up to now almost all commercial semiconductor devices have been fabricated using optical lithography, but for feature dimensions below 180 nm, lithographies based on other forms of radiation are being considered, including x rays, extreme ultraviolet (wavelength = 13 nm), electrons, and ions. Although significant technological difficulties remain, recent advances in ion optics and in mask technology have renewed interest in ion-based lithographic methods. Although focused ion beam (FIB) techniques are too slow for large scale lithography, they may compete with e-beam lithography, e.g., for mask making. Focused ion beams are also widely used in failure
analysis and in circuit restructuring to correct mask errors in new circuit development and they may be useful for other direct, maskless fabrication as discussed below.

A. Masked ion beam lithography

Ion exposure of resist using various proximity stencil masks, as shown in Fig. 61(b), has been explored since the 1970s. Linewidths below 0.1 μm can be exposed and the source of ions (usually protons) is readily available, e.g., from an implanter. The main difficulty has been the production of masks which stop ions in the 100 keV range and at the same time are fabricated at the final dimensions on the sample. Thermal expansion due to heating as well as swelling due to ion bombardment may also need to be considered.

Large area, robust Si membranes with controlled amounts of stress (also used in x-ray lithography) as well as techniques for etching vertical sidewall openings in the mask have been developed. Techniques for controlled radiation cooling of the mask can be used to reduce the undesirable effects of thermal expansion. The advances in ion optics and ion sources permit large areas ~200 mm in diameter to be uniformly irradiated with a divergence of 30 μrad and a controlled telecentricity in the 0.5 mrad range. Thus a large mask-to-wafer gap is possible, e.g., a 1 mm gap would result in a penumbral blur of only 30 nm, which is a key feature for application of this technology. The adjustable telecentricity also permits accurate compensation for mask or wafer expansion.

FIG. 61. Three techniques for producing a patterned dose of ions on a surface: (a) focused ion beam, (b) proximity-masked ion beam, and (c) ion projection lithography.

B. Ion projection lithography

A schematic of an ion projection system is shown in Fig. 62. Early systems have demonstrated the capability of exposing features down to 0.15 μm linewidth, but problems of image distortion and drift were thought to make this technology impractical. An example of a 0.18 μm width line exposed in Ray PN (negative) resist over 0.5 mm steps in the wafer is shown in Fig. 63. Note that there is no change in linewidth in going over the step. Recent improvements have resulted in relatively low distortion over an 8×8 mm field with a 5:1 demagnification of the mask on the wafer and a mechanism for “locking” the pattern to a reference plate above a sample (see Fig. 62). This “beam lock” stabilized the pattern to better than 30 nm against position drift as well as scale drift.

The mask technology for ion projection is easier than for proximity printing that was discussed above. The energy of incident ions is only 10 keV and the dose is spread over a 9× larger area (for 3:1 demagnification). Thus heat input per unit area to the mask is about two orders of magnitude lower. Also, the ion penetration in the mask is much lower (approx 100 nm) and can be limited to a special surface film (e.g., carbon) which limits damage and milling of the Si mem-

FIG. 62. Schematic of ion projection lithography system including pattern lock system to prevent drift in magnification and position (from Ref. 204).
C. Focused ion beams for microfabrication

Focused ion beams provide other methods for pattern modification with extremely high lateral precision that bypass lithography. Highly focused beams can directly mill grooves with widths below 0.1 μm, selectively deposit material from a precursor gas with comparable resolution, or implant dopants into semiconductors. In addition, focused ion beams can increase the resolution of standard materials analysis by SIMS. Scanning ion beam imaging (analogous to scanning electron microscopy but using ions) can be performed, thus permitting all of the microfabrication or analysis to be done at desired, predetermined locations with a high degree of accuracy (better than ±0.1 μm location with respect to alignment masks).

1. Machine design

A focused ion beam system is composed of three parts housed in a vacuum chamber: ion source, ion optical column, and sample stage. The ion source now exclusively used is the liquid metal type (we will discuss an alternative later when we discuss future developments). In this source a needle (often tungsten) is surrounded by molten metal, such as Ga, or alloys such as Au/Si, Au/Si/Be, or Pd/As/B. The needle is on the axis of a concentric extraction electrode and is biased (~10 kV) positively with respect to it. The liquid metal on the end of the needle is pulled into a cusp by the electric field, and ions are emitted from the tip of the cusp. Ga ion sources are used in most of the commercial systems and are quite stable and long lived (1000 or more hours). The ions are emitted into a cone of 20°–30° half-angle at a total current, under optimum conditions of 1–2 μA.

Because the ions from the liquid metal source have an energy spread of at least 5 eV, chromatic aberration limits the useable current in the ion optical column to only a small fraction of the total current. A beam defining aperture at the entrance of the ion optical column restricts the angular divergence to a few mrad and a current that is in the 10 pA–1 nA range. The fundamental factors that limit the beam size are chromatic aberration, spherical aberration, and virtual source size. Practical factors that limit beam size are vibration, electrical noise on the lenses, and stray fields. The lowest beam diameter reported is 8 nm whereas the best commercial systems are now specified to deliver less than 20 nm minimum beam diameter.

To “write” with a focused ion beam over a large area a combination of stage motion and beam deflection is used. Deflecting the beam with a radial electrostatic field produces some blurring of the focal spot due to the energy spread of the ions emitted from the source. The size of the area over which the beam can be deflected depends on the desired minimum beam diameter. Because of source demagnification, smaller beam diameters are achieved at smaller working distances. Thus, both energy spread and finite source size limit the deflection field, i.e., the smaller the desired beam diameter, the smaller the deflection field. In practice, field sizes of order 100–500 μm are used.

2. Applications

The main commercial applications of focused ion beam systems in Si integrated circuit manufacturing are circuit re-wiring, failure analysis, and mask repair. A prototype circuit that fails to operate or needs to be analyzed in sections can be “rewired” with a focused ion beam by milling trenches to cut conductors and by depositing conducting jumpers to...
make new connections (see Fig. 64). In failure analysis, a
trench is cut through a non-working contact, for example, and
the cross section is examined by a scanning electron micro-
scope (SEM) (in some cases built into the same system), by
scanning ion microscopy, or by SIMS. An example of a
cross section cut by a focused ion beam and imaged by scan-
ing ion microscopy is shown in Fig. 65. Mask repair re-
quires metal ion-induced deposition to replace a missing ab-
sorber and ion milling or ion-assisted etching to remove an
unwanted absorber. This becomes particularly challenging
for phase shift masks. These applications depend on three
types of processes: ion milling, ion-assisted etching, i.e.,
milling in the presence of a reactive gas, and ion-induced
deposition from a precursor gas.

a. Ion milling. The special issues that arise when remov-
ing atoms with a focused point beam are redeposition and
milling at non-normal incidence. In general, if the ion
beam is scanned rapidly over an area, then the ion milling
rate is the same as for a broad beam defined over the same
area. However, if the beam is scanned slowly so that each
scan removes a thickness of material that is comparable to
the beam diameter, then locally the ions doing the milling are
not at normal incidence. In addition, milling a deep pit with
a single scan of the ion beam leads to redeposition effects
and a partial filling of the pit by the redeposited material.
This effect has been modeled and the angular distribution of
the patterned substrate material measured. The depend-
ence of milling yield on angle has also been measured for
Ga ions on several materials. As long as milling is done
with rapid, repeated scans, complications can be avoided
and the milling yield is the same as is measured at normal inci-
dence, e.g., ~3 atoms/ion for Si, and ~2 molecules/ion for
SiO2. An example of a cross-sectioned circuit is shown in
Fig. 65.

b. Ion assisted etching. Focused ion beam etching was
reported a number of years ago, and the emergence of
applications where it is needed has recently accelerated its
development. In the repair of photomasks, particularly
phase shift masks, ion milling with no reactive gas leaves a
Ga implant in the quartz substrate which causes an increase
in optical absorption coefficient [especially in the ultraviolet
(UV)]. If a reactive gas is used, the Ga implant density can
be greatly reduced. Also, in the repair of x-ray masks XeF2
gas is used to increase the tungsten removal rate 100X over
milling alone. In addition, the etch is selective and a very
thin underlayer of Cr can be used as an etch stop. Ion-
assisted etching has also been used to increase the removal
rate of SiO2 and Si3N4. This is useful in the rewiring of
integrated circuits and the cross sectioning of integrated cir-
cuits for fault analysis. Highly selective ion-assisted etching
of resist and other organic films has been demonstrated using
H2O as the active gas. This is particularly useful for cleanly
cross sectioning resist profiles, i.e., avoiding redeposition ef-
fects that occur during milling.

c. Ion induced deposition. The deposition of fine patterns
of material using a focused ion beam and a precursor gas is
in some sense the reverse of the etching process. Precursor
gases such as W(CO)6, as well as Cu, Pt, or Au bearing
organometallics, have been used to deposit metals. In addi-
tion, insulators have recently also been deposited. The
metals that are deposited contain up to 50% carbon. This
means that their resistivity is at least one to two orders of
magnitude higher than that of pure metals. Nevertheless,
these “metal” deposits are used to make connections in the
repair and restructuring of integrated circuits. If needed, the
higher resistivity can be compensated for by using a thicker
deposit. In the case of copper and gold deposition the resis-
tivity of the deposit can be reduced to near the value of pure
metal by heating the substrate during deposition to about
100 °C, or by using local laser illumination. An ex-
ample of a rewired IC is shown in Fig. 64. The process
usually consists of milling vias (or ion etching them with a
reactive gas) and then depositing a jumper from one via to
another, as shown in Fig. 64.

FIG. 64. Example of FIB circuit restructuring. A trench is milled to cut a
conductor (left foreground) and a jumper is deposited to join two conductors
(in the center) (from Ref. 211).

FIG. 65. Cross section of a 4 Mbit DRAM imaged with a 30 keV Ga ion
beam. The entire width is 6 µm. The light colored outline across the center
is due to a 7 nm dielectric layer (from Ref. 212).
d. High energy, mass-separated systems. The applications in this area are still in the research phase and include direct maskless, resistless implantation of semiconductors and ion lithography (i.e., resist exposure). The high energy, mass-separated systems operate up to 150 kV accelerating voltage, and provide a variety of ion species, in particular, the common dopants of Si and GaAs. Beam diameters down to 50 nm are available at current densities of 0.1–0.4 A cm$^{-2}$. The sources used are alloys such as Pd/As/B or Au/Si/Be which typically emit both singly and doubly charged ions (for a review, see Ref. 226).

e. Implantation. Two fundamental issues with ion implantation are dose rate effects due to the high instantaneous current density and lateral straggle. The current density is typically three to five orders of magnitude higher than in conventional ion implantation and some dose rate effects have been observed. For example, in GaAs and Si, ions at 260 keV and doses above $10^{14}$ cm$^{-2}$ produce more extensive damage when implanted with a focused ion beam than with a conventional broad beam.\textsuperscript{227} Since the dose in practice is often lower and the scanning can be speeded up to average out the current, deleterious effects of dose rate can be avoided.

Very few measurements of lateral straggle of ions exist. However, with the shrinking of the dimensions of semiconductor devices, lateral straggle may become a practical limitation. Of course, it also limits the minimum dimension with which the density of FIB implanted species inside a semiconductor can be patterned. The lateral straggle has been measured indirectly by implanting gratings of various periods in GaAs using Be and Si ions. For Be ions at 260 keV the lateral straggle appears to be smaller (e.g., <100 nm) than predicted by existing models\textsuperscript{228} (see Fig. 66).

A unique feature of focused ion beam implantation is point-by-point control of the dose. Devices can be fabricated with lateral dose gradients and the dose can be varied from device to device on a chip or a wafer. (For review of this technology, see, for example, Ref. 226.) Some of the unique devices that have been built exploiting the capabilities of focused ion beam include: flash analog/digital (A/D) converters,\textsuperscript{229} tunable Gunn diodes,\textsuperscript{230,231} faster charge coupled devices (CCDs),\textsuperscript{232,233} and enhanced field-effect transistors.\textsuperscript{234,235} An example of the enhanced performance of a 1.6 $\mu$m channel $N$-type MOS (NMOS) transistor is shown in Fig. 67. A line of B, implanted next to the source region of the transistor, increased the transconductance by 30% and increased the output conductance by an order of magnitude.

f. Lithography. At this time there are only two techniques for writing original patterns (as opposed to replicating...
them) at 0.1 μm and below: electron beams and ion beams. (Tunneling microscope-based writing can also be used but is orders of magnitude slower.) Electron beams are at a mature state of development. However, ion beams have an advantage in the absence of the proximity effect and high resist sensitivity, i.e., potentially faster writing speed. Two fundamental issues with ion beam lithography are shot noise (statistical fluctuation in the number of ions per pixel) and the velocity of ions in the column, which limits deflection speed.236

Shot noise simply limits the sensitivity of the resist that can be used, and represents more of a theoretical than practical limitation. In fact, well formed features have been written with as few as 25 ions per pixel²⁰ (in this case 15 × 15 nm, i.e., to form 15-nm-wide lines). For writing 0.1 μm lines a pixel (0.1 × 0.1 μm) will have 100–1000 ions depending on the resist sensitivity.

Beam deflection and blanking are achieved by electrodes adjacent to the beam. A H₂⁺ ion at 100 keV travels 1 cm in 3.2 ns. Since electrodes can be under 1 cm long and the time to expose a 0.075 × 0.075 μm pixel with a 3.6 nA beam is only 26 ns, the travel time for light ions will not be a limitation. This 3.6 nA current is estimated to be the beam current available for a 75 nm beam using the gas field ion source described below. With the development of this source, point beam lithography with ions is projected to be 40–200 times faster than the fastest point e-beam writers demonstrated.²³⁶ An additional advantage is that ion lithography, unlike e-beam lithography, has a negligible proximity effect.

g. Future directions. Recent advances indicate that stable, low energy spread, gas field ion sources may be possible²³⁷ with beam current density in the focal spot in the 80–100 A cm⁻² range, about two orders of magnitude higher than the current density achievable, for example, with Be⁺⁺ ions. Lithography with H₂⁺ ions from a gas field ion source is projected to be 170 times faster for exposure of PMMA and 35 times faster for exposure of negative resist.²³⁶ Since resists are somewhat less sensitive to light ions, shot noise will be less of a limitation, and since light ions have a higher velocity, deflection speeds commensurate with the projected rapid exposure can be achieved. Gas field ion sources also reduce contamination effects that are found with Ga sources.

Although implantation is slow, the focused ion beam can be used to produce devices with a unique performance. In addition to the examples described above, other applications can be envisioned where local implantation is required for special practical functions. For example, some sensors require special ion implants (such as Zn, Na, etc.) in exposed oxide ‘‘gates.’’ Since a given chip may require several such sensors, each implanted in a small area with a different ion species, implanting them with a focused ion beam will eliminate a large number of conventional fabrication steps, and it will be done very rapidly due to the small area. This will, of course, require development of a liquid metal ion source capable of yielding all of the ion species. Examples of 4- or 5-element liquid metal ion sources have been reported.

VII. SUMMARY

This review details the current status of energetic ion beams in the Si integrated circuit industry. Ion implantation continues to be a key element in the processing suite, and this is driving the development of several scientific and technological trends. Paramount among these is the need to understand dopant and defect interactions at the atomic level. As we have shown, considerable progress has been made in unraveling the phenomena underlying transient enhanced diffusion. An important component of this renaissance of defect physics is the use of atomistic modeling which is proving to be an influential tool. It is highly likely that, if this progress continues in atomistic understanding, truly predictive TCAD models for implantation processing will result.

Technological imperatives are demanding implantation at ever lower and ever higher energies. At lower energies, machines have to transport and deliver intense beams in a controlled and monitored fashion—no easy challenge, but progress, as we have described, is being made. A severe experimental challenge for these low energies is the measurement of 2D dopant profiles with lateral resolution of 10 nm. The demand and growth of high energy implantation are increasing rapidly but the current base of machines in place is still small. Limitations here do not appear to depend on machine development but, rather, on the understanding and control of the extended defects in this novel processing regime.

In this review, we have focused somewhat myopically on the Si industry because of the immense interest and scientific richness in Si. Similar questions need to be asked about the role of implantation in compound semiconductors and other electronic and photonic materials. Many of the experimental and theoretical tools developed for Si processing could help advance implantation and defect science in these other materials. Nevertheless, the inherent complexity of the implantation and defect processes in the compound semiconductors may continue to restrict implantation’s widespread use.

Ion implantation was conceived of some forty-four years ago and in the intervening years it has developed into a key element of microelectronic manufacturing such that all integrated circuits are now doped using the technique. It is difficult to imagine the rapid development of integrated circuit technology without implantation. Continuing trends towards smaller size devices and lower thermal budgets are pushing implantation into previously unexplored processing territory. For the future, the challenge for implantation is to determine whether it can be used to dope the very shallow and laterally constrained regions demanded by future device scaling. If implantation cannot be used for doping, what techniques will supplant it?

ACKNOWLEDGMENTS

This panel study was sponsored by the Council on Materials Science of the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences. The members of the panel would like to thank Dr. Jerry Smith of DOE/BES-MS and Professor C. Peter Flynn, University of


Appl. Phys. Rev.: Chason et al. 6557